Signal Processing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics

Toby Christian Lawin-Ore *, Simon Doclo

University of Oldenburg, Department of Medical Physics and Acoustics and Cluster of Excellence "Hearing4All", 26111 Oldenburg, Germany

ARTICLE INFO

Article history:
Received 18 February 2014
Received in revised form
16 June 2014
Accepted 18 June 2014

Keywords:
Multi-channel Wiener filter
Statistical room acoustics
Acoustic sensor network

ABSTRACT

For most multi-microphone noise reduction algorithms, e.g. the multi-channel Wiener filter (MWF), it is well known that the performance depends on the acoustic scenario at hand, i.e. the used microphone array, the position of the desired source and the noise field. Since the position of the desired source is not always known a priori, it is of great interest in many applications to be able to compute the *average performance* for a specific microphone array, which can be obtained by averaging the performance over all feasible source positions. A possible but either time-consuming or computationally complex approach to achieve this is to use measurements or simulations for a large number of source positions.

In this paper, we propose to use the statistical properties of the acoustical transfer functions (ATFs) between the desired source and the microphones to derive *analytical expressions* for the spatially averaged performance measures (output SNR, noise reduction, speech distortion) of the MWF, assuming a homogeneous and known noise field. In addition, we show that although the spatially averaged performance measures do not express the performance of the MWF for a given position of the source and/or the microphones, they can be used to derive approximate analytical expressions for the average performance of the MWF for a given position of the microphones. Experimental results show that the proposed analytical expressions can be used to easily compare the performance of different microphone arrays, e.g. in an acoustic sensor network, without having to measure or numerically simulate a large number of ATFs.

1. Introduction

In many speech communication applications, such as teleconferencing and hearing aids, either a single microphone or a microphone array at a fixed position are typically used to capture the speech signals. As a consequence, the desired source is often located at a large distance from the microphones, possibly resulting in a low input signal-to-noise ratio (SNR) and hence a degraded speech quality. In recent years, research on speech enhancement using spatially distributed microphones has gained significant interest [1–8]. Using spatially distributed microphones or so-called *acoustic sensor networks* (ASNs), the microphones located at distinct places are able to acquire more information about the sound field than a single microphone array at one position, such that the probability that the desired source is close to one of the

* Corresponding author.
E-mail addresses: toby.chris.lawin.ore@uni-oldenburg.de (T.C. Lawin-Ore), simon.doclo@uni-oldenburg.de (S. Doclo).

http://dx.doi.org/10.1016/j.sigpro.2014.06.017
0165-1684/

microphones is higher. For example, ASNs have been considered for applications such as in-car applications [5–7], surveillance [8], teleconferencing [9] and for hearing aid applications [1,2,10–12], where microphone arrays located on different hearing aids (or even other devices) exchange information with each other in order to improve speech intelligibility in noisy environments.

When all microphone signals in an ASN consisting of several spatially distributed microphone arrays are wirelessly transmitted between the different microphone arrays or to a central processing unit, the wireless link would require a large bandwidth. To reduce the required bandwidth of the wireless link, several well-known (centralized) multi-microphone noise reduction algorithms have been extended to the so-called distributed noise reduction algorithms, where each microphone array locally combines its noisy microphone signals and exchanges the resulting output signal with the other microphone arrays in the network in order to estimate a network-wide desired signal. The linearly constrained minimum variance (LCMV) beamformer, which minimizes the noise variance at the output of the beamformer subject to one or more linear constraints (e.g. distortionless response for the desired signal), and the multi-channel Wiener filter (MWF), which minimizes the mean square error (MSE) between the output signal and the reference signal, are two popular classes of multi-microphone noise reduction algorithms [13–16]. Distributed versions of the LCMV beamformer, the minimum variance distortionless response beamformer, which is a special case of the LCMV beamformer, and the generalized sidelobe canceller, which is an alternative implementation of the LCMV beamformer, have been proposed in [3,4,11]. It has to be noted that most algorithms that are based on the LCMV beamformer rely on a priori knowledge or assumptions about the array geometry and the position of the desired source.

Unlike the LCMV beamformer, the MWF does not require the array geometry and the position of the desired source to be known. In the context of ASNs, a distributed MWF (DB-MWF) algorithm has been introduced for binaural speech enhancement, where two hearing aid devices, each having two or more microphones, iteratively exchange locally estimated desired signals [1]. After a few iterations, the DB-MWF converges to the centralized binaural MWF, i.e. the MWF computed using all noisy microphone signals. In [2], a distributed node-specific signal estimation (DANSE) algorithm, which is an extension of the DB-MWF algorithm to more than two microphone arrays and multiple desired sources, has been proposed.

For every noise reduction algorithm it is of significant interest to be able to compute its theoretical performance (e.g. output SNR, noise reduction, speech distortion), which enables us to compare the performance of different microphone arrays [17]. The performance of most multi-microphone noise reduction algorithms obviously depends on the acoustical scenario, i.e. the number and positions of the microphones, the position of the desired source and the noise field. Although being able to compute the performance for a specific position of the desired source and the microphone array is definitely worthwhile, in many applications it is of even greater use to compute the *average performance* for a specific microphone array (e.g. by averaging the performance over all feasible source positions in the room), which enables us to compare the performance of different microphone array topologies. However, computing the performance for a large number of source–microphones' configurations, either requires a large number of acoustic *measurements*, which could be very time-consuming, or the performance needs to be *numerically simulated*, e.g. by simulating the acoustical transfer functions (ATFs) using the image method [18] or room acoustics software, which could be computationally complex. Therefore, it would be very useful to have *analytical expressions* that allow for a faster computation of average performance measures.

In this paper, we only consider the MWF algorithm, which aims to estimate the desired signal component in one of the microphones (referred to as the reference microphone), and we assume that all microphone signals are available on a central processor. In [17], the theoretical performance of the MWF has been analyzed for different noise fields (diffuse and coherent noise sources). It has been shown that the performance (e.g. the output SNR) of the MWF only depends on the noise correlation matrix and the ATFs between the desired source and the microphones. Hence, for every source–microphones' configuration, the theoretical performance can be computed using measured or simulated noise correlation matrices and ATFs.

On the other hand, analytical expressions for spatially averaged performance measures have been derived using *statistical room acoustics* (SRA) for various acoustic signal processing algorithms [19–24]. In [19], a statistical model for the ATFs has been proposed and a method to predict the SNR improvement of a delay-and-sum beamformer with two microphones has been presented. In [20–22], the robustness of single-channel and multi-channel equalization techniques has been analyzed using SRA. Furthermore, in [23] the performance of a blind source separation algorithm has been investigated and in [24] the performance of acoustic crosstalk cancellation has been computed using SRA. Basically, all analytical expressions for the spatially averaged performance measures in the aforementioned methods are based on the statistical ATF model proposed in [25,26], i.e. using the spatial second-order statistics of the ATFs [25–27].

Recently, for a *given relative distance* between the desired source and the microphones and assuming that the noise field is homogeneous and known, spatially averaged performance measures of the MWF have been analytically derived by incorporating the statistical properties of the ATFs into the theoretical expressions for the performance measures of the MWF [28,29]. Simulation results have shown that the spatially averaged performance measures, computed analytically using the statistical properties of ATFs, are similar to the spatially averaged performance measures of the MWF, computed numerically using simulated ATFs. However, it should be realized that the analytical expressions for the spatially averaged performance measures derived in [28,29] do not yet allow us to compute the average performance for a specific microphone array, since only the relative distance between the desired source and the microphones is given.

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

In this paper, we first review the analytical expressions for the spatially averaged performance measures of the MWF, for a given relative distance between the desired source and the microphones, and we then show that for a given position of the microphones the spatially averaged performance measures can be used to derive (approximate) analytical expressions for the average performance of the MWF. The proposed analytical expressions allow for an easy performance comparison of different microphone arrays (with given topologies), without having to measure or numerically simulate ATFs.

This paper is organized as follows. Section 2 describes the notation and the used signal model. In Section 3 the MWF is briefly reviewed and its theoretical performance measures are introduced. In Section 4 the concept of SRA and the statistical properties of ATFs are reviewed. In Section 5 analytical expressions for the spatially averaged performance measures of the MWF, for a given relative distance between the desired source and the microphones, are derived. These analytical expressions are then used to derive analytical expressions for the average performance of the MWF for a given position of the microphones. The validity of all derived analytical expressions is verified by numerical simulations in Section 6 for three different microphone topologies and assuming a diffuse noise field.

2. Notation and signal model

2.1. Notation

Consider the acoustical scenario depicted in Fig. 1 with a single desired source $S(\omega)$ located at position $\mathbf{p}_s=[x_s\ y_s\ z_s]^T$ and M microphones located at positions $\mathbf{p}_m=[x_m\ y_m\ z_m]^T$, $m=0\ldots M-1$. The complete microphone array is described by the $3\times M-$dimensional matrix $\mathbf{P}_{mic}=[\mathbf{p}_0\cdots\mathbf{p}_{M-1}]$, where the topology of the microphone array, i.e., the relative distance between the microphones, is assumed to be fixed but not the location of the microphone array. Since the desired source and the microphone array can be located anywhere in the room, we consider $\mathbf{p}_s$ and $\mathbf{P}_{mic}$ as stochastic variables. We define the stochastic variable $\mathbf{P}=[\mathbf{P}_{mic},\mathbf{p}_s]$ as the combination of the positions of the microphones and the desired source and we define the relative distance between the desired source and the microphones as

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

which is also a stochastic variable. Furthermore, we define the set of all possible realizations of $\mathbf{P}$ in the room as

$$\mathbf{Q}=\{\mathbf{P}^{jk}=[\mathbf{P}^j_{mic},\mathbf{p}^k_s]\ \forall j,k\}, \tag{2}$$

where $\mathbf{P}^j_{mic}$ and $\mathbf{p}^k_s$ represent the jth and kth realization of $\mathbf{P}_{mic}$ and $\mathbf{p}_s$ respectively. We define $\mathbf{Q}^i\subset\mathbf{Q}$ as the subset of realizations with a specific relative distance $\mathbf{d}^i$ between the desired source and the microphones, i.e.,

$$\mathbf{Q}^i=\{\mathbf{P}^{jk}=[\mathbf{P}^j_{mic},\mathbf{p}^k_s]\ \forall j,k|\mathbf{d}^i\}. \tag{3}$$

Moreover, we define the spatial expectation operator $\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{\cdot\}$ as the ensemble average over all realizations of $\mathbf{P}$ with a *given relative distance* $\mathbf{d}^i$ (i.e. over the subset $\mathbf{Q}^i$) and the spatial expectation operator $\mathcal{E}_{\mathbf{P}|\mathbf{P}^j_{mic}}\{\cdot\}$ as the ensemble average over all realizations of $\mathbf{P}$ for a given position $\mathbf{P}^j_{mic}$ of the microphones.

2.2. Signal model

For any realization of the positions of the microphones and the desired source, the microphone signals can be described in the frequency-domain as

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

where $\mathbf{Y}(\omega)=[Y_0(\omega)\cdots Y_{M-1}(\omega)]^T$ denotes the stacked vector of the microphone signals, $\mathbf{H}(\omega)=[H_0(\omega)\cdots H_{M-1}(\omega)]^T$ denotes the stacked vector of the ATFs between the desired speech source $S(\omega)$ and the microphone array, ω is the angular frequency in rad/s 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.,

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

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

$$\mathbf{\Phi}_y=\mathcal{E}\{\mathbf{Y}\mathbf{Y}^H\},\quad \mathbf{\Phi}_x=\mathcal{E}\{\mathbf{X}\mathbf{X}^H\},\quad \mathbf{\Phi}_v=\mathcal{E}\{\mathbf{V}\mathbf{V}^H\}, \tag{6}$$

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

$$\mathbf{\Phi}_y=\mathbf{\Phi}_x+\mathbf{\Phi}_v. \tag{7}$$

Using a robust voice activity detection method, 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 speech pauses.

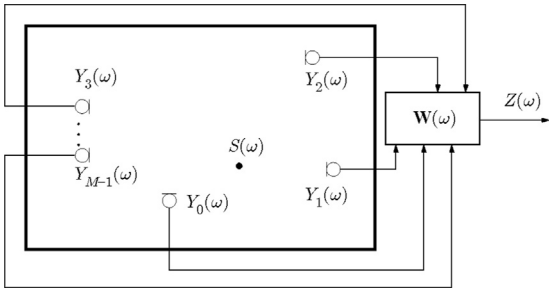

Fig. 1. Acoustic sensor network with M microphones and a single desired source.

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

In the remainder of this paper, a homogeneous noise field[1] is assumed, i.e., the noise component of the microphone signals has the same power spectral density (PSD), i.e. $\phi_v = \mathcal{E}\{|V_m|^2\}$, $m = 1 \ldots M$. Hence, the noise correlation matrix can be expressed as $\mathbf{\Phi}_v = \phi_v \mathbf{\Gamma}_v$, where $\mathbf{\Gamma}_v$ denotes the noise coherence matrix. Furthermore, since a single desired speech source is assumed, the speech correlation matrix $\mathbf{\Phi}_x = \phi_s \mathbf{H}\mathbf{H}^H$ is a rank-one matrix, where ϕ_s represents the PSD of the source S, i.e. $\phi_s = \mathcal{E}\{|S|^2\}$.

3. Multi-channel Wiener filtering

The concept of multi-channel Wiener filtering is based on estimating the speech component X_{m_0} of the m_0th microphone signal, arbitrarily selected as the reference microphone. The MWF produces a minimum-mean-square-error (MMSE) estimate by minimizing the MSE cost function [15,16]

$$\xi(\mathbf{W}) = \mathcal{E}\{|X_{m_0} - \mathbf{W}^H \mathbf{Y}|^2\}. \tag{8}$$

The filter minimizing (8) is given by

$$\mathbf{W}_{m_0} = \mathbf{\Phi}_y^{-1} \mathbf{\Phi}_x \mathbf{e}_{m_0}, \tag{9}$$

where $\mathbf{e}_{m_0}$ is an M-dimensional vector of which the m_0th element is equal to 1 and all other elements are equal to 0, i.e. selecting the column that corresponds to the reference microphone. Using the matrix inversion lemma, it can be shown that (9) can be rewritten as [15]

$$\mathbf{W}_{m_0} = \frac{\mathbf{\Gamma}_v^{-1}\mathbf{H}}{\frac{\phi_v}{\phi_s} + \rho} H_{m_0}^* \tag{10}$$

where H_{m_0} denotes the ATF between the source and the reference microphone,

$$\rho = \mathbf{H}^H \mathbf{\Gamma}_v^{-1} \mathbf{H} \tag{11}$$

and ϕ_s/ϕ_v corresponds to the a priori input SNR.

The (frequency-dependent) *input SNR* of the reference microphone signal is defined as

$$\mathrm{SNR}_{\mathrm{in}} = \frac{\mathcal{E}\{|X_{m_0}|^2\}}{\mathcal{E}\{|V_{m_0}|^2\}} = \frac{\phi_s}{\phi_v} |H_{m_0}|^2. \tag{12}$$

Similar to the input SNR, using (5), (10) and (11), the (frequency-dependent) output SNR of the MWF is defined as

$$\mathrm{SNR}_{out} = \frac{\mathcal{E}\{|Z_x|^2\}}{\mathcal{E}\{|Z_v|^2\}} = \frac{\mathbf{W}_{m_0}^H \mathbf{\Phi}_x \mathbf{W}_{m_0}}{\mathbf{W}_{m_0}^H \mathbf{\Phi}_v \mathbf{W}_{m_0}} = \frac{\phi_s}{\phi_v}\rho. \tag{13}$$

Although the output SNR is commonly used to express the performance of signal enhancement algorithms, it does not show how much noise has been reduced or how much speech has been distorted. The amount of *noise reduction* (NR) can be expressed as

$$\mathrm{NR} = \frac{\mathcal{E}\{|Z_v|^2\}}{\mathcal{E}\{|V_{m_0}|^2\}} = \frac{\mathbf{W}_{m_0}^H \mathbf{\Phi}_v \mathbf{W}_{m_0}}{\phi_v} = \frac{|H_{m_0}|^2 \rho}{\left(\frac{\phi_v}{\phi_s} + \rho\right)^2}, \tag{14}$$

while the amount of *speech distortion* (SD) can be expressed as

$$\mathrm{SD} = \frac{\mathcal{E}\{|Z_x|^2\}}{\mathcal{E}\{|X_{m_0}|^2\}} = \frac{\mathbf{W}_{m_0}^H \mathbf{\Phi}_x \mathbf{W}_{m_0}}{\phi_s |H_{m_0}|^2} = \frac{\rho^2}{\left(\frac{\phi_v}{\phi_s} + \rho\right)^2}. \tag{15}$$

The *SNR improvement* is defined as the ratio of the output SNR and the input SNR at the reference microphone m_0, which can also be expressed as the ratio of the noise reduction and the speech distortion, i.e.,

$$\Delta\mathrm{SNR} = \frac{\mathrm{SNR}_{\mathrm{out}}}{\mathrm{SNR}_{\mathrm{in}}} = \frac{\mathrm{SD}}{\mathrm{NR}} = \frac{\rho}{|H_{m_0}|^2}. \tag{16}$$

Similarly, the MSE of the multi-channel Wiener filter can be computed by inserting (10) into (8), i.e.,

$$\xi(\mathbf{W}_{m_0}) = \mathcal{E}\{|X_{m_0} - \mathbf{W}_{m_0}^H \mathbf{Y}|^2\} = \frac{\phi_v |H_{m_0}|^2}{\left(\frac{\phi_v}{\phi_s} + \rho\right)}. \tag{17}$$

As can be noted from (12)–(17), for a single desired source and a homogeneous noise field, all performance measures of the MWF only depend on the ATF $\mathbf{H}$ between the desired source and the microphones, the spatial characteristics of the noise field described by the noise coherence matrix $\mathbf{\Gamma}_v$ and the a priori input SNR ϕ_s/ϕ_v (except the SNR improvement).

4. Statistical properties of ATFs

In this section, the statistical ATF model proposed in [25] is reviewed. More specifically, the second-order statistics of the direct and the reverberant components of the ATFs are derived, which will be used in Section 5 to compute spatially averaged performance measures of the MWF.

For any realization of the positions of the microphones and the desired source, the sound pressure observed at the mth microphone can be described in the frequency-domain as

$$p_m(\mathbf{P}) = p_{m,d}(\mathbf{P}) + p_{m,r}(\mathbf{P}), \tag{18}$$

where $p_{m,d}(\mathbf{P})$ and $p_{m,r}(\mathbf{P})$ correspond to the direct and the reverberant component, respectively. As shown in [20,30], (18) can be expressed as a function of the ATF, i.e.,

$$p_m(\mathbf{P}) = -j\omega\nu S H_m(\mathbf{P}) = -j\omega\nu S(H_{m,d}(\mathbf{P}) + H_{m,r}(\mathbf{P})), \tag{19}$$

where $H_{m,d}(\mathbf{P})$ and $H_{m,r}(\mathbf{P})$ correspond to the direct and reverberant components respectively of the ATF and ν denotes the density of air. The theory of statistical room acoustics is based on the assumption that the reverberant sound field consists of a large number of plane waves arriving from all directions with randomly distributed amplitudes and phases. Since the reverberant sound pressure is a sum of a large number of independent and identically distributed random variables, the central limit

[1] The assumption of a homogeneous noise field holds for a diffuse noise field and is a good approximation when the microphones are closely spaced.

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

theorem can be applied, and $p_{m,r}(\mathbf{P})$ can be assumed to be zero-mean Gaussian distributed. However, the validity of this assumption only holds if the following conditions are fulfilled [25]:

1. The dimensions of the room should be large relative to the wavelength of the considered signals. This condition is necessary in order to ensure that the average distance between the room resonance frequencies is small enough compared to the mean half-width of the resonances, such that for each frequency a large number of excited room modes are involved in the generation of the reverberant sound field.
2. The considered frequencies should be above the Schroeder frequency, i.e.,

$$f > f_g = 2000\sqrt{T_{60}/V}, \tag{20}$$

where T_{60} is the reverberation time and V is the volume of the room. Under this condition, the number of excited independent room modes is large enough to obtain a Gaussian distribution.
3. The microphones and the source should be located at least half a wavelength away from the walls. For example, for speech signals with a lower frequency of 300 Hz, the microphones and the source should be at least about 0.6 m away from the walls.

When the reverberant sound pressure is zero-mean Gaussian distributed, it can be shown that the spatial correlation between the reverberant sound pressures observed at the mth and the nth microphone can be expressed as [30]

$$\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{p_{m,r}(\mathbf{P})p^*_{n,r}(\mathbf{P})\} = \overline{p}_0^2(\omega)\frac{\sin\left(\frac{\omega}{c}r_{mn}\right)}{\frac{\omega}{c}r_{mn}}, \tag{21}$$

where $r_{mn} = \|\mathbf{p}_m - \mathbf{p}_n\|$ represents the distance between the mth and the nth microphone, c is the speed of sound propagation in air and $\overline{p}_0^2(\omega)$ represents the mean square pressure of the reverberant sound field. The mean square pressure $\overline{p}_0^2(\omega)$ is given by [30]

$$\overline{p}_0^2(\omega) = (\omega\nu)^2\phi_s(\omega)\frac{1-\overline{\alpha}}{\pi\overline{\alpha}A}, \tag{22}$$

where A is the total surface of the walls and $\overline{\alpha} = \sum_n A_n\alpha_n$ is the average absorption coefficient, with A_n and α_n being the surface and the absorption coefficient of the nth wall, respectively. If the reverberation time T_{60} is known, the average absorption coefficient can be approximated using Sabine's formula as [27]

$$\overline{\alpha} = \frac{0.161V}{AT_{60}}. \tag{23}$$

Using (19), (21) and (22), the spatial correlation between the reverberant components of the ATFs can be expressed as

$$\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{H_{m,r}(\mathbf{P})H^*_{n,r}(\mathbf{P})\} = \frac{1-\overline{\alpha}}{\pi\overline{\alpha}A}\frac{\sin\left(\frac{\omega}{c}r_{mn}\right)}{\frac{\omega}{c}r_{mn}} \quad \forall m,n \tag{24}$$

Moreover, given the relative distance $\mathbf{d}^i$ between the source and the microphones, the direct components of the ATFs can be modeled as the free space Green's function, i.e.,

$$H_{m,d}\left(\mathbf{P}|\mathbf{d}^i\right) = \frac{e^{-j(\omega/c)d^i_m}}{4\pi d^i_m} \quad \forall m, \tag{25}$$

where d^i_m is the distance between the source and the mth microphone. Therefore, the spatial correlation between the direct components is a deterministic quantity which only depends on the relative distance between the source and the microphones and is given by

$$\begin{aligned}\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{H_{m,d}(\mathbf{P})H^*_{n,d}(\mathbf{P})\} &= H_{m,d}\left(\mathbf{P}\middle|\mathbf{d}^i\right)H^*_{n,d}\left(\mathbf{P}\middle|\mathbf{d}^i\right)\\ &= \frac{e^{j(\omega/c)(d^i_n - d^i_m)}}{(4\pi)^2 d^i_m d^i_n} \quad \forall m,n\end{aligned} \tag{26}$$

Using the fact that the direct sound pressure $p_{m,d}(\mathbf{P})$ is the same for all realizations $\mathbf{P}^{jk}$ with a *given relative distance* $\mathbf{d}^i$ and using the fact that the reverberant sound pressure is zero-mean Gaussian distributed, the spatial correlation between the direct and reverberant sound pressures is equal to zero, i.e.,

$$\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{p_{m,d}(\mathbf{P})p^*_{n,r}(\mathbf{P})\} = 0 \quad \forall m,n. \tag{27}$$

Hence, using (19) and (25), the direct and reverberant components of the ATFs are spatially uncorrelated, i.e.,

$$\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{H_{m,d}(\mathbf{P})H^*_{n,r}(\mathbf{P})\} = 0 \quad \forall m,n \tag{28}$$

Finally, using (24), (26), and (28), the spatial expectation of the energy density spectrum of the mth ATF can be expressed as

$$\begin{aligned}\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{|H_m(\mathbf{P})|^2\} &= \mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{|H_{m,d}(\mathbf{P})|^2\} + \mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{|H_{m,r}(\mathbf{P})|^2\}\\ &= \frac{1}{(4\pi d^i_m)^2} + \frac{1-\overline{\alpha}}{\pi\overline{\alpha}A} \quad \forall m\end{aligned} \tag{29}$$

As can be observed, the spatial expectation of the energy density spectrum only depends on the distance d^i_m between the desired source and the mth microphone and on the room properties $(A, \overline{\alpha})$.

5. Spatially averaged performance measures of MWF

Using the spatial correlation properties of the ATFs derived in the previous section, analytical expressions for the spatially averaged performance measures of the MWF, for a given relative distance $\mathbf{d}^i$ between the desired source and the microphones will be derived in Section 5.1. These analytical expressions will then be used in Section 5.2 to derive (approximate) analytical expressions for the average performance of the MWF for a given position $\mathbf{P}^j_{mic}$ of the microphones.

5.1. Spatially averaged performance of MWF for a given relative distance $\mathbf{d}^i$

The objective of this section is to incorporate the statistical properties of the ATFs derived in Section 4 into the performance measures of the MWF derived in Section

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

3 and to derive analytical expressions for the spatially averaged performance measures for a given relative distance $\mathbf{d}^i$ between the desired source and the microphones.

Without loss of generality, we define $\widetilde{\mathrm{PM}}(\mathbf{d}^i)$ as a spatially averaged performance measure for a given relative distance $\mathbf{d}^i$, i.e.,

$$\widetilde{\mathrm{PM}}(\mathbf{d}^i) = \mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{\mathrm{PM}(\mathbf{P})\} \tag{30}$$

where PM represents either ρ, $\mathrm{SNR}_{\mathrm{in}}$, $\mathrm{SNR}_{\mathrm{out}}$, NR, SD, ΔSNR or ξ defined in Section 3. It is of utmost importance to realize that $\widetilde{\mathrm{PM}}(\mathbf{d}^i)$ denotes the performance averaged over all realizations $\mathbf{P}^{jk} \in \mathbf{Q}^i$ (i.e. for a given relative distance $\mathbf{d}^i$), but is not equal to the performance for each realization in this subset, i.e.

$$\mathrm{PM}(\mathbf{P}^{jk}|\mathbf{d}^i) \neq \widetilde{\mathrm{PM}}(\mathbf{d}^i) \quad \forall j,k. \tag{31}$$

This is due to the fact that for the computation of $\widetilde{\mathrm{PM}}(\mathbf{d}^i)$ neither the location of the microphone array nor the position of the desired source is fixed.

Using the fact that the ATFs can be decomposed into direct and reverberant components, the factor ρ in (11) can be rewritten as

$$\begin{aligned}\rho(\mathbf{P}) = {} & \mathbf{H}_d^H(\mathbf{P})\mathbf{\Gamma}_v^{-1}\mathbf{H}_d(\mathbf{P}) + \mathbf{H}_d^H(\mathbf{P})\mathbf{\Gamma}_v^{-1}\mathbf{H}_r(\mathbf{P}) \\ & + \mathbf{H}_r^H(\mathbf{P})\mathbf{\Gamma}_v^{-1}\mathbf{H}_d(\mathbf{P}) + \mathbf{H}_r^H(\mathbf{P})\mathbf{\Gamma}_v^{-1}\mathbf{H}_r(\mathbf{P}),\end{aligned} \tag{32}$$

where $\mathbf{H}_d(\mathbf{P})$ and $\mathbf{H}_r(\mathbf{P})$ correspond to the direct and the reverberant component of the ATFs. Without loss of generality, $\mathbf{H}_1^H(\mathbf{P})\mathbf{\Gamma}_v^{-1}\mathbf{H}_2(\mathbf{P})$ can be expressed as

$$\mathbf{H}_1^H(\mathbf{P})\mathbf{\Gamma}_v^{-1}\mathbf{H}_2(\mathbf{P}) = \sum_{m=1}^{M}\sum_{n=1}^{M} \check{\gamma}_{mn} H_{m,1}^*(\mathbf{P}) H_{n,2}(\mathbf{P}), \tag{33}$$

where $\mathbf{H}_1(\mathbf{P})$ and $\mathbf{H}_2(\mathbf{P})$ can represent either $\mathbf{H}_d(\mathbf{P})$ or $\mathbf{H}_r(\mathbf{P})$ and $\check{\gamma}_{mn}$ denotes the coefficients of the inverse noise coherence matrix $\mathbf{\Gamma}_v^{-1}$. Hence, $\rho(\mathbf{P})$ can be written as

$$\begin{aligned}\rho(\mathbf{P}) = \sum_{m=1}^{M}\sum_{n=1}^{M} \check{\gamma}_{mn} & (H_{m,d}^*(\mathbf{P})H_{n,d}(\mathbf{P}) + H_{m,d}^*(\mathbf{P})H_{n,r}(\mathbf{P}) \\ & + H_{m,r}^*(\mathbf{P})H_{n,d}(\mathbf{P}) + H_{m,r}^*(\mathbf{P})H_{n,r}(\mathbf{P})).\end{aligned} \tag{34}$$

Using (28), the spatially averaged value of ρ for a *given relative distance* $\mathbf{d}^i$ is then equal to

$$\widetilde{\rho}(\mathbf{d}^i) = \sum_{m=1}^{M}\sum_{n=1}^{M} \check{\gamma}_{mn}(\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{H_{m,d}^*(\mathbf{P})H_{n,d}(\mathbf{P})\} + \mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{H_{m,r}^*(\mathbf{P})H_{n,r}(\mathbf{P})\}), \tag{35}$$

which, using (24) and (26) is equal to

$$\widetilde{\rho}\left(\mathbf{d}^i\right) = \sum_{m=1}^{M}\sum_{n=1}^{M} \check{\gamma}_{mn}\left(\frac{e^{\mathrm{j}(\omega/c)(d_n^i - d_m^i)}}{(4\pi)^2 d_m^i d_n^i} + \frac{1-\overline{\alpha}}{\pi\overline{\alpha}A}\frac{\sin\left(\frac{\omega}{c}r_{mn}\right)}{\frac{\omega}{c}r_{mn}}\right) \tag{36}$$

and only depends on the relative distance between the source and the microphones, the room properties $(A, \overline{\alpha})$, the noise coherence matrix and the microphone array topology. Analytical expressions for several spatially averaged performance measures of the MWF for a *given relative distance* $\mathbf{d}^i$ will now be derived.

Using (12) and (29), the spatially averaged *input* SNR can be analytically expressed as

$$\widetilde{\mathrm{SNR}}_{\mathrm{in}}\left(\mathbf{d}^i\right) = \frac{\phi_s}{\phi_v}\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{|H_{m_0}(\mathbf{P})|^2\} = \frac{\phi_s}{\phi_v}\left(\frac{1}{(4\pi d_{m_0}^i)^2} + \frac{1-\overline{\alpha}}{\pi\overline{\alpha}A}\right) \tag{37}$$

Using (13) and (36), the spatially averaged *output SNR* can be analytically expressed as

$$\widetilde{\mathrm{SNR}}_{\mathrm{out}}\left(\mathbf{d}^i\right) = \frac{\phi_s}{\phi_v}\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{\rho(\mathbf{P})\} = \frac{\phi_s}{\phi_v}\widetilde{\rho}\left(\mathbf{d}^i\right) \tag{38}$$

While analytical expressions for the spatially averaged input SNR and output SNR can be derived without any approximation, approximations are required in order to derive similar expressions for the spatially averaged noise reduction, speech distortion, SNR improvement and MSE.

Using (14), the spatially averaged *noise reduction* is given by

$$\widetilde{\mathrm{NR}}\left(\mathbf{d}^i\right) = \mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\left\{\frac{|H_{m_0}(\mathbf{P})|^2\rho(\mathbf{P})}{\left(\frac{\phi_v}{\phi_s} + \rho(\mathbf{P})\right)^2}\right\}. \tag{39}$$

To compute the expected value of a function of two random variables $\rho(\mathbf{P})$ and $|H_{m_0}(\mathbf{P})|^2$, we propose to use an approximation based on the first-order Taylor expansion. If the higher-order derivatives can be neglected at the expansion point, the expected value of a function of two random variables can be approximated by the function of the expected value of the two random variables (cf. Appendix A). Although the first-order Taylor expansion might not be a good approximation for all functions, this approximation will be validated by the experimental results in Section 6.2. The spatially averaged noise reduction can then be approximated as

$$\begin{aligned}\widetilde{\mathrm{NR}}\left(\mathbf{d}^i\right) \approx \frac{\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{|H_{m_0}(\mathbf{P})|^2\}\widetilde{\rho}(\mathbf{d}^i)}{\left(\frac{\phi_v}{\phi_s} + \widetilde{\rho}\left(\mathbf{d}^i\right)\right)^2} = {} & \left(\frac{1}{(4\pi d_{m_0}^i)^2} + \frac{1-\overline{\alpha}}{\pi\overline{\alpha}A}\right) \\ & \frac{\widetilde{\rho}(\mathbf{d}^i)}{\left(\frac{\phi_v}{\phi_s} + \widetilde{\rho}\left(\mathbf{d}^i\right)\right)^2}\end{aligned} \tag{40}$$

Similarly, using (15)–(17) and their first-order Taylor expansion, the spatially averaged speech distortion, the spatially averaged SNR improvement and the spatially averaged mean square error can be approximated as

$$\widetilde{\mathrm{SD}}\left(\mathbf{d}^i\right) = \mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\left\{\frac{\rho^2(\mathbf{P})}{\left(\frac{\phi_v}{\phi_s} + \rho(\mathbf{P})\right)^2}\right\} \approx \frac{\widetilde{\rho}^2(\mathbf{d}^i)}{\left(\frac{\phi_v}{\phi_s} + \widetilde{\rho}\left(\mathbf{d}^i\right)\right)^2} \tag{41}$$

$$\begin{aligned}\Delta\widetilde{\mathrm{SNR}}(\mathbf{d}^i) &= \mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\left\{\frac{\rho(\mathbf{P})}{|H_{m_0}(\mathbf{P})|^2}\right\} \approx \frac{\widetilde{\rho}(\mathbf{d}^i)}{\mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{|H_{m_0}(\mathbf{P})|^2\}} \\ &= \frac{\widetilde{\mathrm{SNR}}_{\mathrm{out}}(\mathbf{d}^i)}{\widetilde{\mathrm{SNR}}_{\mathrm{in}}(\mathbf{d}^i)} = \frac{\widetilde{\mathrm{SD}}(\mathbf{d}^i)}{\widetilde{\mathrm{NR}}(\mathbf{d}^i)}\end{aligned} \tag{42}$$

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

$$\widetilde{\xi}\left(\mathbf{d}^i\right) \approx \frac{\phi_v \mathcal{E}_{\mathbf{P}|\mathbf{d}^i}\{|H_{m_0}(\mathbf{P})|^2\}}{\frac{\phi_v}{\phi_s} + \widetilde{\rho}\left(\mathbf{d}^i\right)} \tag{43}$$

Again, as can be observed from Eqs. (37)–(43), all derived spatially averaged performance measures of the MWF only depend on the distance between the desired source and the microphones, the room properties, the noise coherence matrix and the microphone array topology.

5.2. Average performance of MWF for a given position $\mathbf{P}^j_{mic}$

In the previous section, analytical expressions for the spatially averaged performance measures have been derived for a *given relative distance* between the source and the microphones, i.e. neither the location of the microphone array nor the position of the source is fixed. As a more useful performance measure enabling to e.g. compare the performance of different microphone topologies, we would actually like to derive analytical expressions for the average performance of the MWF for a given position $\mathbf{P}^j_{mic}$ of the microphones, i.e.

$$\widetilde{\mathrm{PM}}(\mathbf{P}^j_{mic}) = \mathcal{E}_{\mathbf{P}|\mathbf{P}^j_{mic}}\{\mathrm{PM}(\mathbf{P})\} \tag{44}$$

where PM again represents either ρ, $\mathrm{SNR}_{\mathrm{in}}$, $\mathrm{SNR}_{\mathrm{out}}$, NR, SD, ΔSNR or ξ. However, note that it is not straightforward to derive analytical expressions for the average performance measures of the MWF similarly as in Section 5.1, since to the best of our knowledge no analytical expressions for the spatial correlations $\mathcal{E}_{\mathbf{P}|\mathbf{P}^j_{mic}}\{H_{m,d}(\mathbf{P})H^*_{n,d}(\mathbf{P})\}$, $\mathcal{E}_{\mathbf{P}|\mathbf{P}^j_{mic}}\{H_{m,r}(\mathbf{P})H^*_{n,r}(\mathbf{P})\}$ and $\mathcal{E}_{\mathbf{P}|\mathbf{P}^j_{mic}}\{H_{m,d}(\mathbf{P})H^*_{n,r}(\mathbf{P})\}$ can be computed using statistical room acoustics. Nevertheless, we will show that using the spatially averaged performance measures $\widetilde{\mathrm{PM}}(\mathbf{d}^i)$ approximate analytical expressions for the average performance measures $\widetilde{\mathrm{PM}}(\mathbf{P}^j_{mic})$ can be derived.

Remembering that the stochastic variable $\mathbf{P}$ is a combination of the positions of the microphones and the source, the average performance measure in (44) can be written as

$$\begin{aligned}\mathcal{E}_{\mathbf{P}|\mathbf{P}^j_{mic}}\{\mathrm{PM}(\mathbf{P})\} &= \mathcal{E}_{\mathbf{p}_s}\{\mathrm{PM}([\mathbf{P}^j_{mic}, \mathbf{p}_s])\} \\ &= \int \mathrm{PM}([\mathbf{P}^j_{mic}, \mathbf{p}_s]) f_{\mathbf{p}_s}(\mathbf{p}_s)\, d\mathbf{p}_s,\end{aligned} \tag{45}$$

where $f_{\mathbf{p}_s}(\mathbf{p}_s)$ denotes the probability density function of the source position $\mathbf{p}_s$. For the derivation, we assume free-field conditions where the positions of the desired source $\mathbf{p}_s$ are uniformly distributed inside a sphere centered around the microphone array. Although we realize that these assumptions are quite unrealistic (due to room reflections and the typically non-spherical shape of a room), the simulation results in Section 6 show that the derived expressions provide a good approximation for realistic reverberant rooms. Now consider two different orientations $\mathbf{P}^1_{mic}$ and $\mathbf{P}^2_{mic}$ of the microphone array (both in the center of the sphere). For any source position $\mathbf{p}^1_s$ inside the sphere, there always exists a corresponding source position $\mathbf{p}^2_s$ such that for a homogeneous noise field, the performance of the MWF for both combinations of the orientations of the microphone array and the source positions is equal, i.e.,

$$\mathrm{PM}([\mathbf{P}^1_{mic}, \mathbf{p}^1_s]) = \mathrm{PM}([\mathbf{P}^2_{mic}, \mathbf{p}^2_s]). \tag{46}$$

Since the source position is assumed to be uniformly distributed, the average performance measures over all possible positions of the desired source for both orientations of the microphone array are also equal, i.e.,

$$\begin{aligned}\mathcal{E}_{\mathbf{P}|\mathbf{P}^1_{mic}}\{\mathrm{PM}(\mathbf{P})\} &= \int \mathrm{PM}([\mathbf{P}^1_{mic}, \mathbf{p}_s]) f_{\mathbf{p}_s}(\mathbf{p}_s)\, d\mathbf{p}_s \\ &= \int \mathrm{PM}([\mathbf{P}^2_{mic}, \mathbf{p}_s]) f_{\mathbf{p}_s}(\mathbf{p}_s)\, d\mathbf{p}_s = \mathcal{E}_{\mathbf{P}|\mathbf{P}^2_{mic}}\{\mathrm{PM}(\mathbf{P})\}.\end{aligned} \tag{47}$$

Assuming furthermore that all realizations of $\mathbf{P}_{mic}$ (with a fixed microphone array topology) can be considered as different orientations of the microphone array,[2] the average performance is equal for all realizations, such that

$$\widetilde{\mathrm{PM}}(\mathbf{P}^j_{mic}) = \mathcal{E}_{\mathbf{P}_{mic}}\{\mathcal{E}_{\mathbf{P}|\mathbf{P}_{mic}}\{\mathrm{PM}(\mathbf{P})\}\} \quad \forall j. \tag{48}$$

This will be verified by simulations in Section 6.3 and it should be realized that although the average performance is assumed to be independent of the location of the microphone array it obviously still depends on the topology of the microphone array. Using the law of total expectation [31], i.e.

$$\mathcal{E}_{\mathbf{P}}\{\mathrm{PM}(\mathbf{P})\} = \mathcal{E}_{\mathbf{P}_{mic}}\{\mathcal{E}_{\mathbf{P}|\mathbf{P}_{mic}}\{\mathrm{PM}(\mathbf{P})\}\} = \mathcal{E}_{\mathbf{d}}\{\mathcal{E}_{\mathbf{P}|\mathbf{d}}\{\mathrm{PM}(\mathbf{P})\}\}, \tag{49}$$

the average performance can be computed as

$$\widetilde{\mathrm{PM}}(\mathbf{P}^j_{mic}) = \mathcal{E}_{\mathbf{d}}\{\mathcal{E}_{\mathbf{P}|\mathbf{d}}\{\mathrm{PM}(\mathbf{P})\}\} = \int \mathcal{E}_{\mathbf{P}|\mathbf{d}}\{\mathrm{PM}(\mathbf{P})\} f_{\mathbf{d}}(\mathbf{d})\, d\mathbf{d} \tag{50}$$

with $f_{\mathbf{d}}(\mathbf{d})$ denoting the probability density function of the relative distance $\mathbf{d}$ between the source and the microphones. Solving this multi-dimensional integral by inserting either (37), (38), and (40)–(42) or (43) into (50) is a tedious problem. However, this integral can be approximated by a finite Riemann sum (e.g. assuming a uniform distribution for the relative distance $\mathbf{d}$) as

$$\widetilde{\mathrm{PM}}\left(\mathbf{P}^j_{mic}\right) \approx \frac{1}{N_d} \sum_{i=1}^{N_d} \widetilde{\mathrm{PM}}\left(\mathbf{d}^i\right) \tag{51}$$

where N_d is the total number of considered relative distances. By plugging in any of the spatially averaged performance measures for a given relative distance derived in Section 5.1 into (51), the average performance measure for a given position of the microphones, i.e. actually for a given topology of the microphone array, can be computed.

6. Simulation results

In order to validate the analytical expressions derived in the previous sections, we now present simulation results. The experimental setup is described in Section 6.1. In Section 6.2, the analytical expressions for the spatially averaged performance measures derived in Section 5.1 are compared with simulated spatially averaged performance

[2] This corresponds to assuming an infinitely large sphere around the microphone array

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

measures, numerically computed using simulated ATFs. In Section 6.3, the validity of the assumptions in Section 5.2 is verified. In Section 6.4, the analytically computed average performance measures are compared with numerically simulated average performance measures for different microphone arrays.

6.1. Experimental setup

In a room with dimensions 8 m × 6 m × 5 m, resulting in a volume $V = 240\ \mathrm{m}^3$ and a total wall surface $A = 236\ \mathrm{m}^2$, we consider the acoustic sensor network depicted in Fig. 2 with 3 nodes, where each node consists of 4 microphones with an inter-microphone distance of 4 cm. The performance will be evaluated for three different microphone arrays with different topologies. For the first topology the first node is selected ($M=4$ microphones), for the second topology the first and second nodes are selected ($M=8$ microphones) and for the third topology all nodes are selected ($M=12$ microphones). Two different reverberation times T_{60} will be considered, i.e. 0.4 s and 0.8 s (resulting in a Schroeder frequency f_g in (20) of 82 Hz and 116 Hz). For each realization of the positions of the desired source and the microphones, room impulse responses have been simulated using the image model [18,32], and the corresponding ATFs have been calculated. The length of the simulated room impulse responses is $L=4096$ samples and the sampling frequency $f_s = 16{,}000$ Hz. For all experiments, a diffuse noise field has been assumed and the noise coherence matrix was theoretically computed using

$$\gamma_{mn}(\omega) = \frac{\sin\left(\frac{\omega}{c} r_{mn}\right)}{\frac{\omega}{c} r_{mn}}, \tag{52}$$

where $\gamma_{mn}(\omega)$ represents the coefficients of the noise coherence matrix $\mathbf{\Gamma}_v(\omega)$ and the speed of sound propagation in air $c = 340$ m/s. Without loss of generality, the a

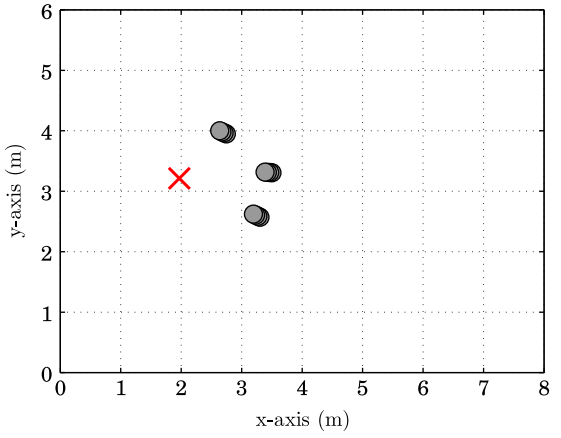

Fig. 2. Acoustic sensor network with 3 nodes.

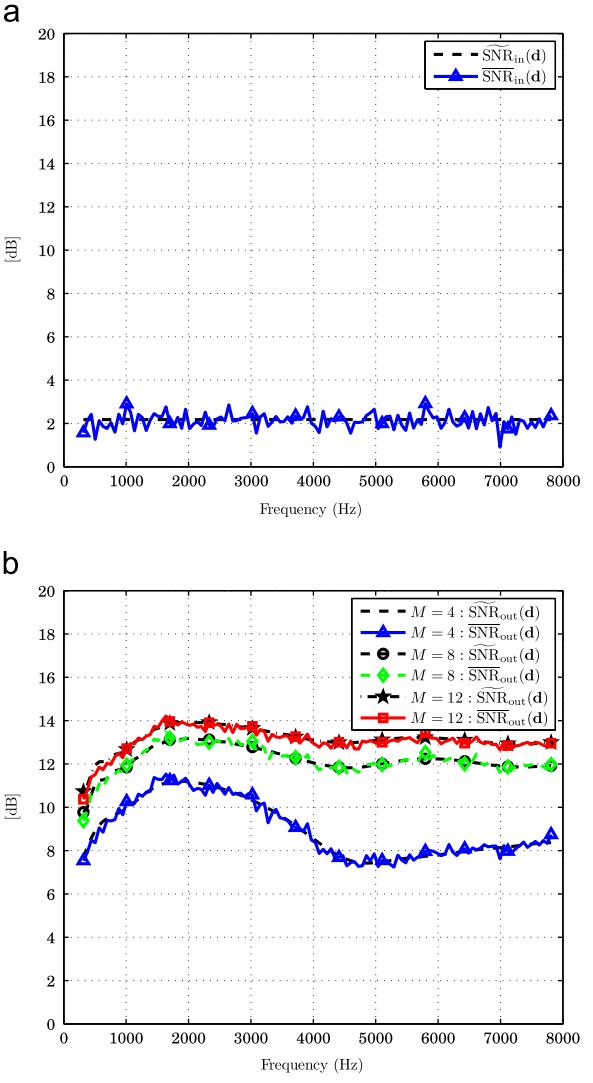

Fig. 3. Simulated spatially averaged performance of MWF using $N=1000$ realizations and analytical results obtained using statistical room acoustics: (a) input SNR, (b) output SNR.

Table 1
Scenarios for source–microphones configurations.

Microphone array (MA)	M	Relative distance (m)
MA_1	4	$\mathbf{d} = [1.39\ 1.43\ 1.47\ 1.51]^T$
MA_2	8	$\mathbf{d} = [1.39\ 1.43\ 1.47\ 1.51\ 1.08\ 1.09\ 1.10\ 1.11]^T$
MA_3	12	$\mathbf{d} = [1.39\ 1.43\ 1.47\ 1.51\ 1.08\ 1.09\ 1.10\ 1.11\ 2.13\ 2.17\ 2.20\ 2.24]^T$

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

priori input SNR ϕ_s/ϕ_v is assumed to be frequency-independent. Furthermore, for all experiments, we select the first microphone of the first node as the reference microphone of the MWF, i.e., $m_0 = 1$.

6.2. Spatially averaged performance measures for a given $\mathbf{d}$

In this section, the analytical expressions for the spatially averaged performance measures $\widetilde{\mathrm{PM}}(\mathbf{d})$, for a *given relative distance* $\mathbf{d}$ between the desired source and the microphones (derived in Section 5.1), are compared to simulated spatially averaged performance measures $\overline{\mathrm{PM}}(\mathbf{d})$, which can be numerically computed as

$$\overline{\mathrm{PM}}(\mathbf{d}) = \frac{1}{N}\sum_{j,k} \mathrm{PM}\left(\left[\mathbf{P}^j_{mic}, \mathbf{p}^k_s\right] | \mathbf{d}\right), \tag{53}$$

where N represents the total number of realizations of the positions of the source and the microphones, and PM represents either $\mathrm{SNR}_{\mathrm{in}}$, $\mathrm{SNR}_{\mathrm{out}}$, NR, SD, ΔSNR or ξ. We have used $N = 1000$ and the different realizations $\mathbf{P}^{jk}$ have been generated by rotating and translating the source–microphones configuration, keeping the relative distance $\mathbf{d}$ constant and considering only the realizations that are located within the room and half a wavelength away from the walls. For the considered microphone array topologies, three different source–microphones configurations have been used (cf. Table 1) and for the specific realization depicted in Fig. 2 also the position of the source has been indicated (cross-marker). In this experiment, we have used a reverberation time $T_{60} = 0.4$ s, resulting in an average absorption coefficient $\overline{\alpha} \approx 0.40$.

Figs. 3–5 compare the simulated spatially averaged performance measures $\overline{\mathrm{SNR}}_{\mathrm{in}}(\mathbf{d})$, $\overline{\mathrm{SNR}}_{\mathrm{out}}(\mathbf{d})$, $\overline{\mathrm{NR}}(\mathbf{d})$, $\overline{\mathrm{SD}}(\mathbf{d})$, $\overline{\Delta\mathrm{SNR}}(\mathbf{d})$, and $\overline{\xi}(\mathbf{d})$, numerically computed using simulated ATFs, with the spatially averaged performance measures $\widetilde{\mathrm{SNR}}_{\mathrm{in}}(\mathbf{d})$, $\widetilde{\mathrm{SNR}}_{\mathrm{out}}(\mathbf{d})$, $\widetilde{\mathrm{NR}}(\mathbf{d})$, $\widetilde{\mathrm{SD}}(\mathbf{d})$, $\widetilde{\Delta\mathrm{SNR}}(\mathbf{d})$, and $\widetilde{\xi}(\mathbf{d})$, calculated using the analytical expressions derived in

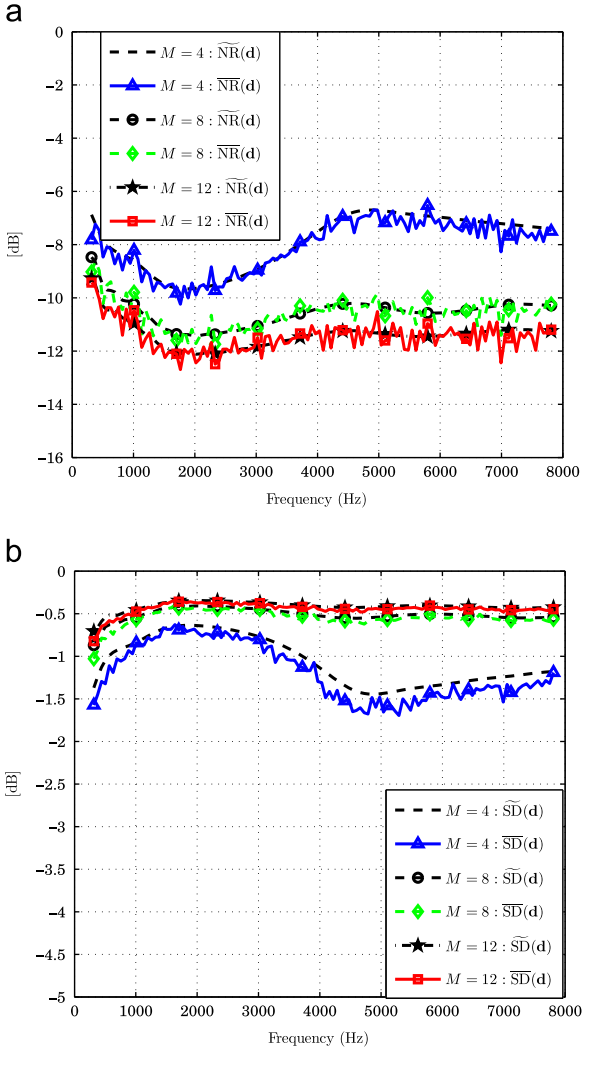

Fig. 4. Simulated spatially averaged performance of MWF using $N = 1000$ realizations and analytical results obtained using statistical room acoustics: (a) noise reduction, (b) speech distortion.

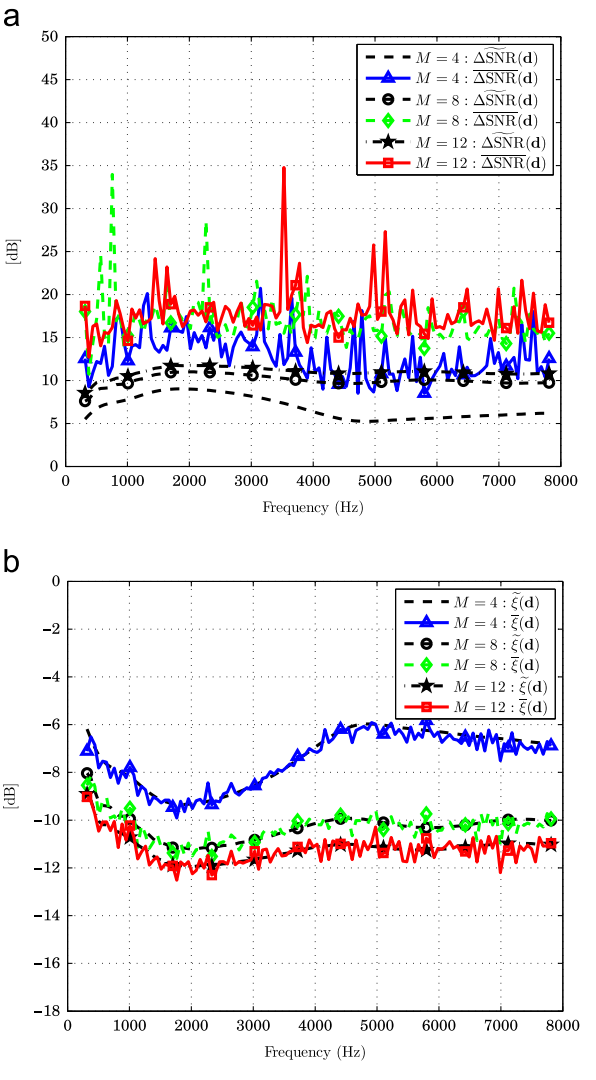

Fig. 5. Simulated spatially averaged performance of MWF using $N = 1000$ realizations and analytical results obtained using statistical room acoustics: (a) SNR improvement, (b) MSE.

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

Section 5.1. Since the first node is part of all considered microphone arrays, the spatially averaged input SNR is the same for all microphone array topologies. Therefore, in order to avoid overcrowded plots, the spatially averaged input SNR $\widetilde{\mathrm{SNR}}_{\mathrm{in}}(\mathbf{d})$ in Fig. 3(a) is shown only for the microphone array with $M=4$ microphones. As can be observed from these figures, on one hand the analytically computed spatially averaged input SNR, output SNR, noise reduction, speech distortion and minimum MSE correspond very well to the numerically simulated spatially averaged performance measures, for all considered microphone arrays and for the complete frequency range. This shows that the first-order Taylor expansion used for deriving analytical expressions for the spatially averaged noise reduction, speech distortion and minimum MSE in Section 5.1 is a good approximation. Therefore, if the relative distance between the source and the microphones and the room properties (A, $\overline{\alpha}$) are known and if the noise coherence matrix is given, the statistical properties of the ATFs can be used to analytically compute the spatially averaged input SNR, output SNR, noise reduction, speech distortion and minimum MSE of the MWF. On the other hand, as can be seen from Fig. 5(a), there is a substantial deviation between the analytically computed spatially averaged SNR improvement and the numerically simulated spatially averaged SNR improvement. This is most likely due to the fact that for some realizations $\mathbf{P}^{jk}$ the magnitude $|H_{m_0}|$ of the ATF is very small (i.e. close to 0) for some frequencies, such that the numerically simulated spatially averaged SNR improvement using (53) is biased.

Fig. 6 shows the (broadband) root mean square error (RMSE) between the spatially averaged performance measures, calculated using the analytical expressions, and the spatially averaged performance measures, numerically computed using simulated ATFs, as a function of the number of realizations N in (53). The RMSE for each performance measure is calculated as

$$\mathrm{RMSE}_{\mathrm{PM}}(N)=\sqrt{\sum_{\omega}|\widetilde{\mathrm{PM}}(\mathbf{d})-\overline{\mathrm{PM}}(\mathbf{d})|^2}. \tag{54}$$

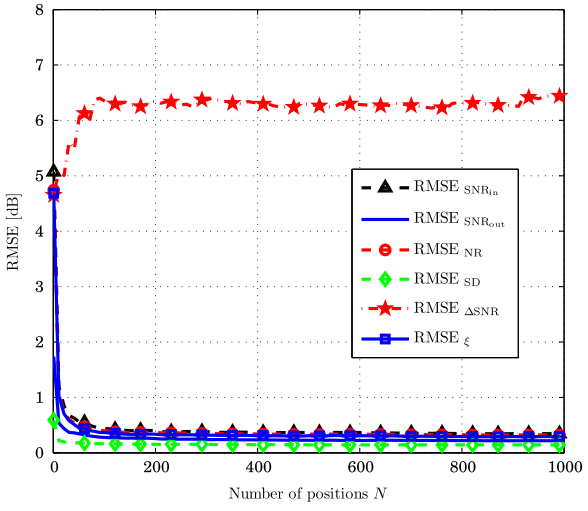

Fig. 6. Root mean square error between numerically simulated and analytical results obtained using statistical room acoustics (microphone array 1).

As can be seen in Fig. 6, the spatially averaged performance measures are not equal to the performance measures of the MWF for a single specific position of the desired source and the microphones. Moreover, the larger the number of realizations N, the smaller the RMSE for all performance measures (except for the SNR improvement). For a large number of realizations, the RMSE of the spatially averaged performance measures (except the SNR improvement) converges to nearly zero, showing the good estimation accuracy of the derived analytical expressions for the spatially averaged performance measures calculated using the analytical expressions. The fact that the RMSEs do not converge exactly to zero may be explained by imperfections of the image model or the assumptions and approximations used in Section 5.1.

6.3. *Dependency of the average performance measures on the location of the microphone array*

In this section, we would like to verify using simulations the crucial assumption in Eq. (48) that the average performance measures of the MWF are independent of the location of the microphone array with a certain topology. In this experiment, we have used a reverberation time $T_{60}=0.4$ s and the microphone array with $M=4$ microphones has been placed at 100 different locations in the room. For each location of the microphone array, the average performance measures have been numerically computed as

$$\overline{\mathrm{PM}}(\mathbf{P}_{mic})=\frac{1}{N_s}\sum_{k=1}^{N_s}\mathrm{PM}\left(\mathbf{P}_{mic},\mathbf{p}_s^k\right), \tag{55}$$

where N_s represents the total number of realizations of the source position ($N_s=2000$).

Fig. 7 shows the average performance at frequency $f=1890$ Hz for different positions of the microphones $\mathbf{P}_{mic}$, i.e. different locations of the microphone array. As can be observed, the performance is fairly constant for different locations of the microphone array with standard deviations in the range of 0.10–0.75 dB. These variations are due

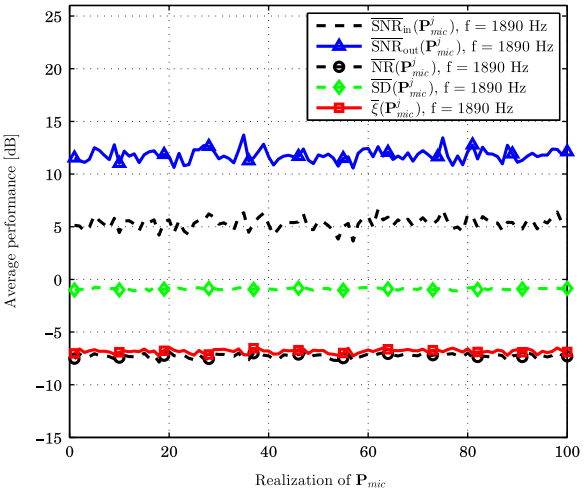

Fig. 7. Average performance measures for different positions of the microphones $\mathbf{P}_{mic}$, i.e. different locations of the microphone array ($M=4$).

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

to the fact that a reverberant rectangular room has been used instead of the sphere with free-field conditions assumed in Section 5.2. Similar results are obtained for other frequencies, microphone topologies and reverberation times.

6.4. Average performance measures for different microphone topologies

In this section, the analytical expressions for the average performance measures for a given position $\mathbf{P}_{mic}$ of the microphones with a certain topology (derived in Section 5.2) are compared with the numerically simulated average performance measures using (55). For the sake of completeness, the average SNR improvement has also been considered although it was shown in Section 6.2 that the numerically simulated spatially averaged SNR improvement $\overline{\Delta\mathrm{SNR}}(\mathbf{d})$ does not correspond to the analytically computed spatially averaged SNR improvement $\Delta\widetilde{\mathrm{SNR}}(\mathbf{d})$.

In this experiment, we consider the same microphone array topologies as in Section 6.2 and two different reverberation times, i.e. $T_{60}=0.4$ s, and $T_{60}=0.8$ s. For computing the average performance measures using (51), a total number of relative distances $N_d=2000$ have been used. Figs. 8–10 compare the numerically simulated average performance measures, i.e., $\overline{\mathrm{SNR}}_{\mathrm{in}}(\mathbf{P}_{mic})$, $\overline{\mathrm{SNR}}_{\mathrm{out}}(\mathbf{P}_{mic})$, $\overline{\mathrm{NR}}(\mathbf{P}_{mic})$, $\overline{\mathrm{SD}}(\mathbf{P}_{mic})$, $\overline{\Delta\mathrm{SNR}}(\mathbf{P}_{mic})$, $\overline{\xi}(\mathbf{P}_{mic})$ with the analytical expressions $\widetilde{\mathrm{SNR}}_{\mathrm{in}}(\mathbf{P}_{mic})$, $\widetilde{\mathrm{SNR}}_{\mathrm{out}}(\mathbf{P}_{mic})$, $\widetilde{\mathrm{NR}}(\mathbf{P}_{mic})$, $\widetilde{\mathrm{SD}}(\mathbf{P}_{mic})$, $\Delta\widetilde{\mathrm{SNR}}(\mathbf{P}_{mic})$, and $\widetilde{\xi}(\mathbf{P}_{mic})$, calculated using (51), for the three considered microphone array topologies. As can be observed, all numerically simulated average performance measures (except for the SNR improvement) correspond well to the average performance measures calculated using the analytical expressions, which only require the topology of the microphone array and the room properties to be known. Similar results are obtained for other

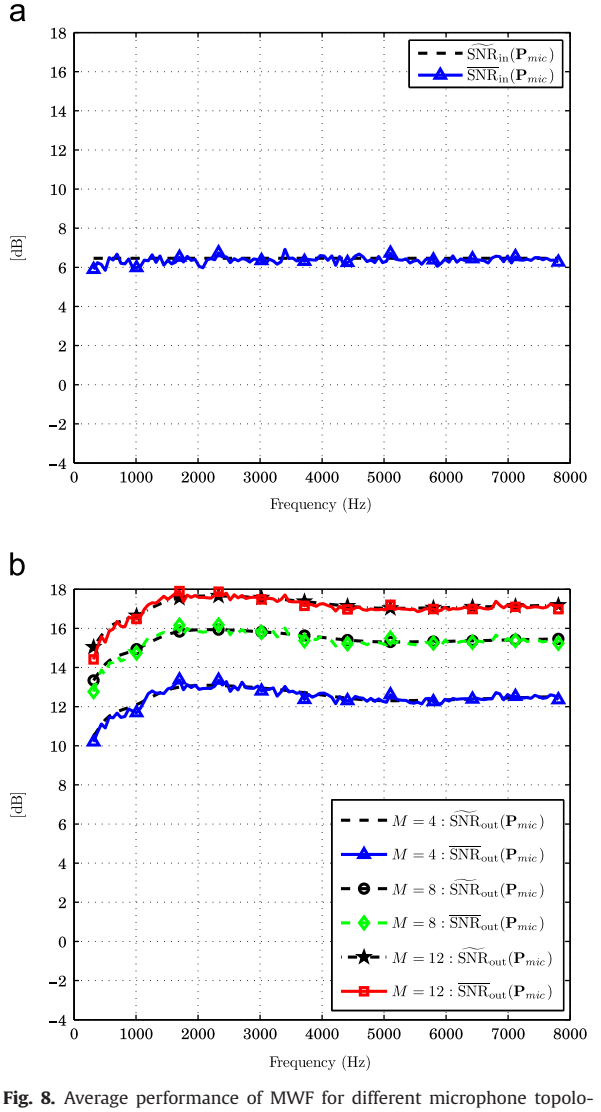

Fig. 8. Average performance of MWF for different microphone topologies: (a) input SNR, (b) output SNR ($T_{60}=0.4$ s).

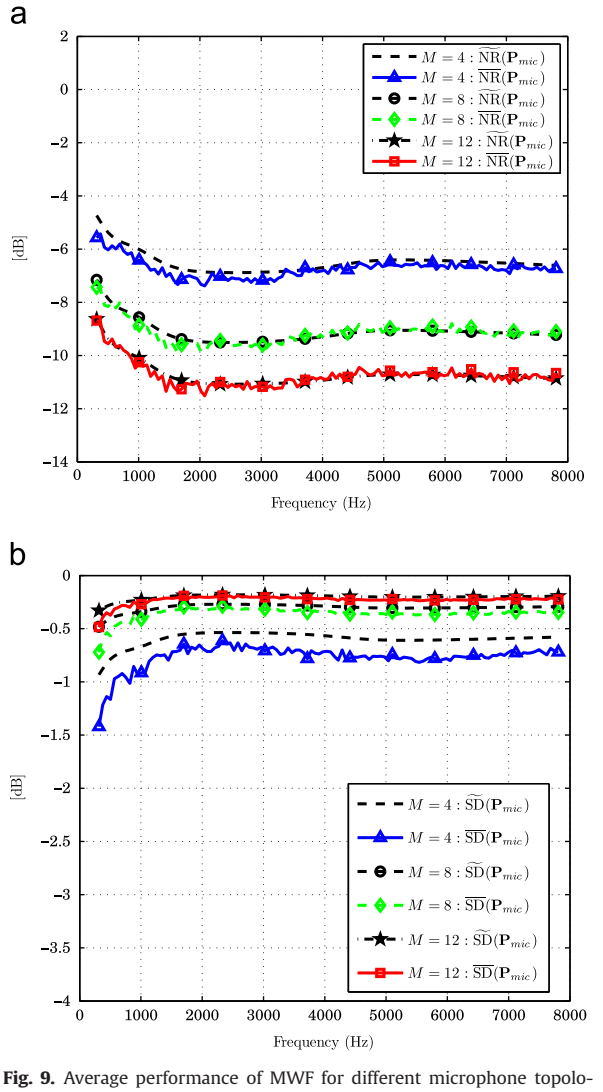

Fig. 9. Average performance of MWF for different microphone topologies: (a) noise reduction, (b) speech distortion ($T_{60}=0.4$ s).

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

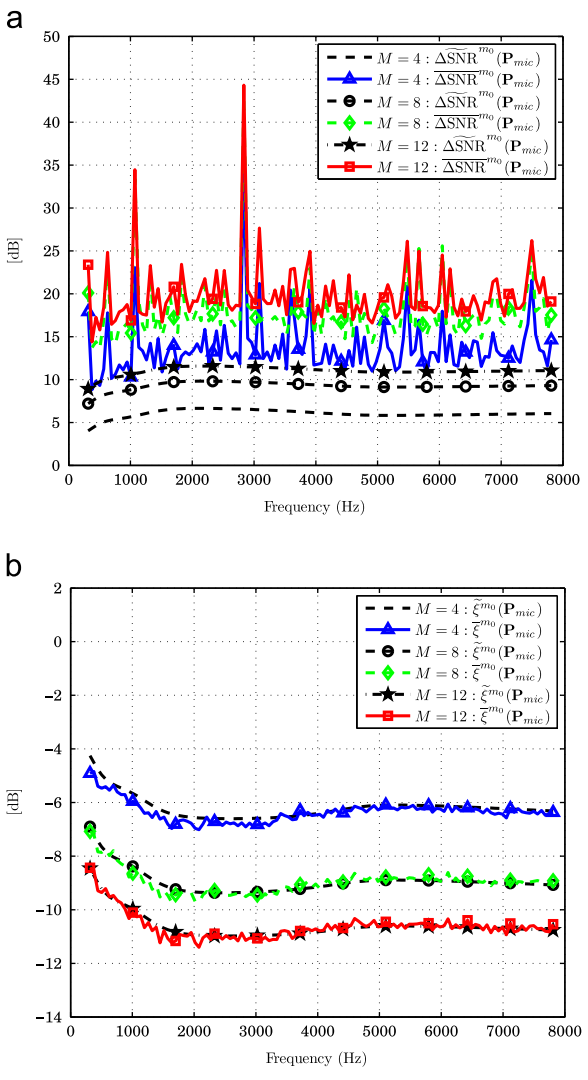

Fig. 10. Average performance of MWF for different microphone topologies: (a) SNR improvement, (b) MSE ($T_{60} = 0.4$ s).

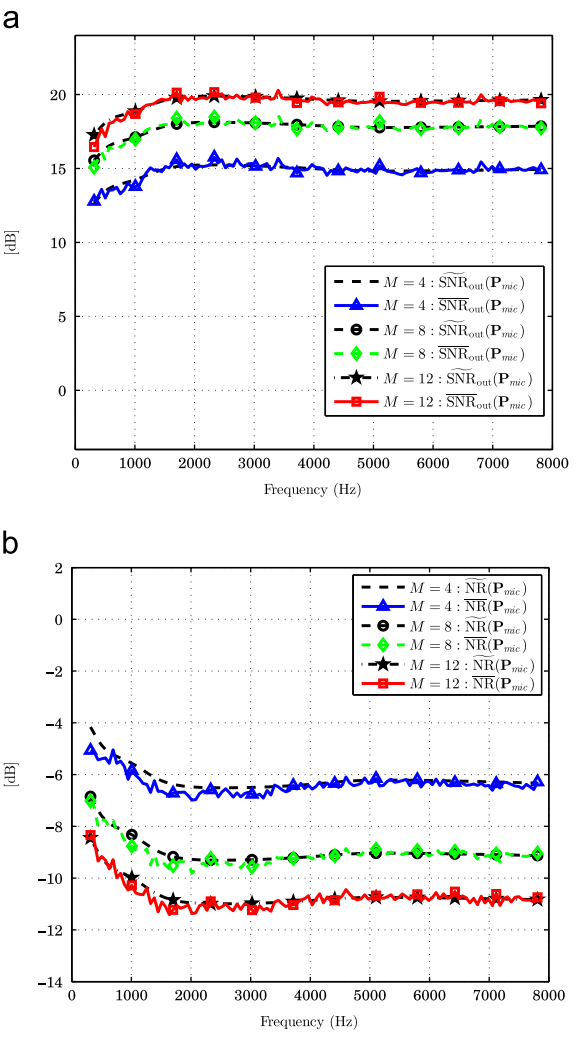

Fig. 11. Average performance of MWF for different microphone topologies: (a) output SNR, (b) noise reduction ($T_{60} = 0.8$ s).

reverberation times. For example, Fig. 11 compares the analytically calculated average output SNR and noise reduction with the numerically simulated average output SNR and noise reduction for $T_{60} = 0.8$ s.

In addition, all presented results in Figs. 8, 9, and 11 clearly show the relation between the average performance measures of the MWF and the number of microphones in a diffuse noise field. For example, as expected, the larger the number of microphones, the higher the average output SNR and the smaller the average speech distortion. Therefore, the analytically computed average performance measures can be used to compare the performance of different microphone arrays without having to measure or simulate the ATFs.

7. Conclusion

In this paper, analytical expressions for the spatially averaged performance measures of the MWF for a *given relative distance* between the desired source and the microphones have been derived by incorporating the statistical properties of the ATFs into the theoretical formulas for the performance of the MWF in a homogeneous noise field. The derived analytical expressions only depend on the room properties (dimensions, reverberation time) and the distance between the source and the microphones. Despite the fact that the analytical expressions for the spatially averaged performance measures for a given relative distance correspond well to the numerically simulated spatially averaged performance measures, they do not directly enable to us compute the average performance of the MWF for a specific position of the microphones. However, in addition we have shown that the spatially averaged performance measures of the MWF can be used to derive a good approximation for the average performance measures given the position of the microphones, i.e. for a given location of the microphone array with a certain topology. Simulation results for several microphone array topologies and reverberation times have shown that these analytical approximations are similar to the results obtained using simulated ATFs, providing an efficient way to

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017

compare the performance of different microphone array topologies, e.g. in an acoustic sensor network, without having to measure or numerically simulate the ATFs.

Acknowledgments

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).

Appendix A. First-order Taylor expansion

Consider two random variables X and Y with $\mu_x = \mathcal{E}\{X\}$ and $\mu_y = \mathcal{E}\{Y\}$. The Taylor expansion of a differentiable function $f(x,y)$ around (μ_x, μ_y) is given by

$$f(x,y) = f(\mu_x,\mu_y) + f'_x(\mu_x,\mu_y)(x-\mu_x) + f'_y(\mu_x,\mu_y)(y-\mu_y) + \hat{f}(x,y), \tag{A.1}$$

where f'_x and f'_y represent the first-order partial derivative with respect to x and y, respectively and $\hat{f}(x,y)$ represents a function of the higher-order partial derivatives of $f(x,y)$. Assuming that all partial derivatives, except the first-order partial derivatives, can be neglected at the expansion point (μ_x, μ_y), then $f(x,y)$ can be approximated by the first-order Taylor expansion, i.e.,

$$f(x,y) \approx f(\mu_x,\mu_y) + f'_x(\mu_x,\mu_y)(x-\mu_x) + f'_y(\mu_x,\mu_y)(y-\mu_y). \tag{A.2}$$

Taking the expectation of both sides of the approximated Taylor expansion yields

$$\mathcal{E}\{f(x,y)\} \approx f(\mu_x,\mu_y). \tag{A.3}$$

References

[1] S. Doclo, T. van den Bogaert, J. Wouters, M. Moonen, Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids, IEEE Trans. Audio Speech Lang. Process. 17 (1) (2009) 38–51.
[2] A. Bertrand, M. Moonen, Distributed adaptive node-specific signal estimation in fully connected sensor networks-part I: sequential node updating, IEEE Trans. Signal Process. 58 (10) (2010) 5257–5291.
[3] A. Bertrand, M. Moonen, Distributed LCMV beamforming in a wireless sensor network with single-channel per-node signal transmission, IEEE Trans. Signal Process. 61 (13) (2013) 3447–3459.
[4] S.M. Golan, S. Gannot, I. Cohen, Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks, IEEE Trans. Audio Speech Lang. Process. 21 (2) (2013) 343–356.
[5] J. Freudenberger, S. Stenzel, B. Venditti, Microphone diversity combining for in-car applications, EURASIP J. Adv. Signal Process., 2010, article ID 509541.
[6] S. Stenzel, J. Freudenberger, Blind matched filtering for speech enhancement with distributed microphones, J. Electr. Comput. Eng., 2012, article ID 169853.
[7] T. Matheja, M. Buck, T. Fingscheidt, A dynamic multi-channel speech enhancement system for distributed microphones in a car environment, EURASIP J. Adv. Signal Process. 2013 (2013) 191.
[8] S.M. Golan, S. Gannot, I. Cohen, Performance of the SDW-MWF with randomly located microphones in a reverberant enclosure, IEEE Trans. Audio Speech Lang. Process. 21 (7) (2013) 1513–1523.
[9] S. Srinivasan, Using a remote wireless microphone for speech enhancement in non-stationary noise, in: Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Prague, Czech Republic, 2011, pp. 4641–4644.
[10] A. Bertrand, M. Moonen, Robust distributed noise reduction in hearing aids with external acoustic sensor nodes, EURASIP J. Adv. Signal Process., 2009, article ID 530435.
[11] S.M. Golan, S. Gannot, I. Cohen, A reduced bandwidth binaural MVDR beamformer, in: Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC), Tel-Aviv, Israel, 2010, pp. 145–148.
[12] T.C. Lawin-Ore, S. Doclo, Analysis of rate constraints for MWF-based noise reduction in acoustic sensor networks, in: Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Prague, Czech Republic, 2011, pp. 269–272.
[13] B. van Veen, K. Buckley, Beamforming: a versatile approach to spatial filtering, IEEE ASSP Mag. 5 (2) (1988) 4–24.
[14] S. Gannot, I. Cohen, Adaptive beamforming and postfiltering, in: Springer Handbook of Speech Processing, Part H, Springer, Berlin, Heidelberg, 2008, pp. 945–978 (Chapter 47).
[15] S. Doclo, S. Gannot, M. Moonen, A. Spriet, Acoustic beamforming for hearing aid applications, in: Handbook on Array Processing and Sensor Networks, Wiley, 2010, pp. 269–302 (Chapter 9).
[16] S. Doclo, A. Spriet, J. Wouters, M. Moonen, Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction, Speech Commun. Spec. Issue Speech Enhanc. 49 (7–8) (2007) 636–656.
[17] A. Spriet, M. Moonen, J. Wouters, Robustness analysis of multichannel Wiener filtering and Generalized Sidelobe Cancellation for multi-microphone noise reduction in hearing aid applications, IEEE Trans. Speech Audio Process. 13 (4) (2005) 487–503.
[18] J. Allen, D. Berkley, Image method for efficiently simulating small-room acoustics, J. Acoust. Soc. Am. 65 (1979) 943–950.
[19] M. Kompis, N. Dillier, Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. I. Prediction of the signal-to-noise-ratio improvement, J. Acoust. Soc. Am. 109 (3) (2001) 1123–1133.
[20] B.D. Radlovic, R.C. Williamson, R.A. Kennedy, Equalization in an acoustic reverberant environment: robustness results, IEEE Trans. Speech Audio Process. 8 (3) (2000) 311–319.
[21] F. Talantzis, D.B. Ward, Robustness of multichannel equalization in an acoustic reverberant environment, J. Acoust. Soc. Am. 114 (2) (2003) 833–841.
[22] S. Bharitkar, P. Hilmes, C. Kyriakakis, Robustness of spatial average equalization: a statistical reverberation model approach, J. Acoust. Soc. Am. 116 (2004) 3491–3497.
[23] F. Talantzis, D.B. Ward, P.A. Naylor, Expected performance of a family of blind source separation algorithms in a reverberant room, in: Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Montreal, Canada, 2004, pp. 61–64.
[24] D.B. Ward, On the performance of acoustic crosstalk cancellation in a reverberant environment, J. Acoust. Soc. Am. 110 (2) (2001) 1195–1198.
[25] M.R. Schroeder, Statistical parameters of the frequency response curves of large rooms, J. Audio Eng. Soc. Am. 35 (5) (1987) 299–306.
[26] M.R. Schroeder, Frequency correlation functions of frequency responses in rooms, J. Acoust. Soc. Am. 34 (12) (1962) 1819–1823.
[27] H. Kuttruff, Room Acoustics, fifth edition, Spon press, London and New York, 2009.
[28] T.C. Lawin-Ore, S. Doclo, Using statistical room acoustics for analysing the output SNR of the MWF in acoustic sensor networks, in: Proceedings of European Signal Processing Conference (EUSIPCO), Bucharest, Romania, 2012, pp. 1259–1263.
[29] T.C. Lawin-Ore, S. Doclo, Using statistical room acoustics for computing the spatially averaged performance of the multichannel Wiener filter based noise reduction, in: Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC), Aachen, Germany, 2012, pp. 145–148.
[30] P.M. Morse, K.U. Ingard, Theoretical Acoustics, McGraw-Hill, London, Boston, 1968.
[31] N.A. Weiss, A Course in Probability, Addison Wesley, Boston, 2005.
[32] E.A.P. Habets, Room impulse response (RIR) generator, available: ⟨http://home.tiscali.nl/ehabets/rirgenerator.html⟩.

Please cite this article as: T.C. Lawin-Ore, S. Doclo, Analysis of the average performance of the multi-channel Wiener filter for distributed microphone arrays using statistical room acoustics, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.06.017