Performance of the SDW-MWF With Randomly Located Microphones in a Reverberant Enclosure

Shmulik Markovich-Golan, *Member, IEEE*, Sharon Gannot, *Senior Member, IEEE*, and Israel Cohen, *Senior Member, IEEE*

***Abstract*—Beamforming with wireless acoustic sensor networks (WASNs) has recently drawn the attention of the research community. As the number of microphones grows it is difficult, and in some applications impossible, to determine their layout beforehand. A common practice in analyzing the expected performance is to utilize statistical considerations. In the current contribution, we consider applying the speech distortion weighted multi-channel Wiener filter (SDW-MWF) to enhance a desired source propagating in a reverberant enclosure where the microphones are randomly located with a uniform distribution. Two noise fields are considered, namely, multiple coherent interference signals and a diffuse sound field. Utilizing the statistics of the acoustic transfer function (ATF), we derive a statistical model for two important criteria of the beamformer (BF): the signal to interference ratio (SIR), and the white noise gain. Moreover, we propose reliability functions, which determine the probability of the SIR and white noise gain to exceed a predefined level. We verify the proposed model with an extensive simulative study.**

***Index Terms*—Optimal filtering, beamforming, performance bounds, room acoustics.**

I. Introduction

TECHNOLOGY advances in recent years in the fields of nano-technology, micro electro-mechanic systems (MEMS) and communication have paved the road for utilizing wireless sensor networks (WSNs) in a variety of applications. A WSN is a wireless network of nodes, distributed over a wide area. Each node, comprises of sensors, actuators, a processing unit and a communication transceiver. The goal of the WSN is to perceive some physical phenomenon, analyze it, and yield an estimated parameter or enhanced signal. The advantages of the WSN over the classic fusion center architecture, in which all sensors are physically connected to a single processing unit, are in its scalability, robustness to failures and higher performance. Relaxing the limitation of all sensors to be wired to the central processing unit, allows for larger area spread of the sensors, thus enabling better coverage, a closer distance to the phenomenon origin and hence better signal to noise ratio (SNR) in the signal acquisition process. Alongside its many advantages, the WSN architecture holds some great challenges too. These challenges result from the limited resources that each node has, namely, battery, processing power and communication-bandwidth. Estrin *et al.* [1] and Culler *et al.* [2] survey the topic of WSNs.

In the current contribution, the WASN, in which the sensors are microphones and the physical phenomena are the propagating speech and noise sources, is discussed. Over the past years extensive research has been made on the subject [3]–[5]. For a survey on distributed beamforming algorithms for WASN please refer to Bertrand [6].

We consider the single desired speaker contaminated by either multiple coherent interference signals or a diffuse sound field. Several criteria for designing the BF and enhancing the desired signal exist (see Van Veen and Buckley [7] for a comprehensive survey). Here, the SDW-MWF beamformer design criterion [8] is adopted. The SDW-MWF distinguishes between two contributions to the mean squared error (MSE), namely, desired signal distortion and residual noise. By introducing a weight μ, a better control of the tradeoff between these two origins of error is obtained. Two special cases of the SDW-MWF are the multi-channel Wiener filter (MWF), which achieves the minimal mean squared error (MMSE), and the minimum variance distortionless response (MVDR), which minimizes the noise variance while maintaining a desired response towards the desired speaker. Spriet *et al.* analyze the performance of the SDW-MWF [9].

As nodes of a WSN are spread out in large quantities, it is difficult to control their exact positions and design a desired beampattern as in classical BF applications. Considering electro-magnetic signals for communication, Lo [10] proposed to apply a statistical model to the sensors locations and to analyze the statistics of the beampattern. He considered the case of a random linear array and a single desired source in the far-field regime. Lo derived analytical expressions for the properties of the beampattern designed using the data-independent delay and sum beamformer (DS-BF). The DS-BF delays the various microphone signals such that the desired speaker component is coherently aligned and sums them. A statistical model for the directivity gain, the beam-width and the sidelobe level were derived. Analytical expressions for the general distribution of the sensors locations are derived, as well as specific expressions for the uniform distribution. Ochiai *et al.* [11] extended the analysis to nodes that are randomly located in a two dimensional disc with a uniform distribution. They considered the application of collaborative beamforming (CB), in which each node has an omni-directional antenna and a group, or a cluster

Manuscript received August 18, 2012; revised December 15, 2012, March 09, 2013; accepted March 12, 2013. Date of publication March 27, 2013; date of current version April 15, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lauri Savioja.

S. Markovich-Golan and S. Gannot are with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan 56000, Israel (e-mail: shmulik.markovich@gmail.com; sharon.gannot@biu.ac.il).

I. Cohen is with the Department of Electrical Engineering, The Technion, Haifa 32000, Israel (e-mail: icohen@ee.technion.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2013.2255280

of nodes, wishes to transmit a signal to some target node. By properly designing the phases of the transmitted signals, the nodes are able to construct a beam aimed at the target node. Ahmed and Vorobyov [12], further extended the discussion, and considered the CB application with three dimensional clusters of randomly located nodes with a normal distribution. They derived analytical expressions for the sidelobe level, and an upper bound for its outage probability. The outage probability of the sidelobe levels is defined as the probability that the maximal sidelobe level is higher than a desired level. Huang *et al.* [13] generalize the discussion regarding the nodes distribution and propose a unified analysis for the properties of the beampattern in the CB application with arbitrary node position statistics. Hardwick *et al.* [14] propose a simple model for the beam-width of a two dimensional cluster of nodes. They considered multiple concurrent transmission channels in a WSN and gave an approximation for the total communication throughput. Kerby and Bernhard [15] considered a periodic array comprising of a common basic sub-array of randomly located nodes. They characterize the probabilistic behavior of the beampattern in several periodic constellations.

In the field of WASN, Jan and Flanagan [16] considered applying the matched filter array (MFA) to a two dimensional randomly located array of microphones. They experimentally showed the superiority of the MFA over the DS-BF in extracting sound sources, while keeping low sidelobe levels. Kodrasi *et al.* [17] show that the array performance depends significantly on the microphone locations and compared various heuristic array design optimization methods. They considered the superdirective BF, which maximizes the directivity index. In [18] we considered the data dependent MWF, which is constructed according to the specific desired speech statistics as well as noise statistics. Assuming a linear array of randomly located microphones with a uniform distribution, we derived analytical expressions for the statistics of the SIR and its reliability, i.e., the probability that the SIR will exceed a desired level. The case-study considered was a desired speaker and a coherent noise source in the far-field regime of a non-reverberant environment.

Considering acoustic signals propagation, Schroeder, in the 1950s, proposed a stochastic model for the room impulse response (RIR) and the respective frequency correlation between microphones [19]. This work was further developed by Polack [20]. In [21] Schroeder investigated the frequency correlation between frequency responses in a room. For a survey on the topic please refer to Kuttruff [22] and to Jot *et al.* [23].

In this contribution we extend [18] and consider an array of randomly located microphones with a uniform distribution in a reverberant enclosure. A single desired speaker is assumed. Utilizing the statistical model of the RIR, we derive a statistical model for the SIR and white noise gain and introduce their reliability functions for the SDW-MWF in several noise scenarios. Specifically, we consider the case of $P < M$ coherent noise sources, where M is the number of microphones, and the case of a diffuse sound field, an infinite number of uncorrelated sources arriving from all directions. The derived statistical model, and the reliability functions can be used to determine the number of microphones needed to assure a desired performance level (with a controlled level of uncertainty). In the derivations we model the ATF as a sum of a direct arrival component and a reverberant component (comprising high order reflections). Note that this is an approximated model, since the early reflections are not considered.

The structure of the paper is as follows. The problem is formulated in Section II. We briefly present the SDW-MWF, and derive expressions for SIR and white noise gain in Section III. In Section IV we derive a statistical model for the ATF, and experimentally verify its validity. Section V is dedicated to deriving a statistical model for the BF criteria, namely SIR and white noise gain, in various noise fields, namely, few coherent noises $P < M$ and diffuse sound field. In Section VI we compare the proposed model and the empirical results obtained from an extensive simulative study. The work is summarized and conclusions are drawn Section VII.

II. Problem Formulation

In the current contribution we analyze the performance of the SDW-MWF, aiming to enhance a desired source contaminated by interference sources. We consider the case of randomly located microphones with a uniform distribution in a reverberant enclosure, and derive a statistical model for the performance.

Consider a $D_x \times D_y \times D_z$ dimensional reverberant room, in which M microphones are randomly located with a uniform distribution. The microphone locations are given in a Cartesian coordinate system, with the origin at the center of the room:

$$\mathbf{r}^m \triangleq [\, r_x^m \quad r_y^m \quad r_z^m \,]^T \tag{1}$$

for $m = 1, \ldots, M$, where $(\bullet)^T$ denotes the transpose operator. Throughout the paper, the short time Fourier transform (STFT) domain is considered, where ℓ and k denote the frame and frequency indices. Let $s_d(\ell, k)$ be a desired speaker positioned at $\mathbf{r}_d$ in the enclosure. The signals received by the microphones are given by:

$$\mathbf{z}(\ell, k) \triangleq \mathbf{h}_d(\ell, k) s_d(\ell, k) + \mathbf{v}(\ell, k) \tag{2}$$

where $\mathbf{h}_d(\ell, k) \triangleq [h_d^1(\ell, k) \; \cdots \; h_d^M(\ell, k)]^T$ denotes the ATF relating the desired speech signal and the microphones, and $\mathbf{v}(\ell, k)$ is a vector comprised of the interference signals. At this point, a general noise field is assumed. In the next sections several specific cases are addressed, explicitly, $P < M$ coherent interference sources as well as a diffuse sound field are considered. Denote the received signals, and the interference covariance matrices, respectively, as:

$$\begin{aligned} \mathbf{\Phi}_{zz}(\ell, k) &\triangleq \mathrm{E}\{\mathbf{z}(\ell, k)\mathbf{z}^\dagger(\ell, k)\} \\ &= \delta_d(\ell, k)\mathbf{h}_d(\ell, k)\mathbf{h}_d^\dagger(\ell, k) + \mathbf{\Phi}_{vv}(\ell, k) \quad (3a) \\ \mathbf{\Phi}_{vv}(\ell, k) &\triangleq \mathrm{E}\{\mathbf{v}(\ell, k)\mathbf{v}^\dagger(\ell, k)\} \quad (3b) \end{aligned}$$

where $\mathrm{E}\{\bullet\}$ denotes the expectation operator, $(\bullet)^\dagger$ denotes the transpose-conjugate operator, and $\delta_d(\ell, k) \triangleq \mathrm{E}\{|s_d(\ell, k)|^2\}$ denotes the power spectral density (PSD) of the desired signal. We assume that the desired speech signal and the interference signals are statistically independent. For brevity, hereafter the frame index ℓ is omitted from the covariance matrices and the ATFs of the desired and interfering sources. The frequency index $k = 1, \ldots, K$ where K is the window length, is also omitted, and the subsequent derivations should be understood as frequency dependent.

In the following section, the SDW-MWF is briefly presented, and its performance criteria are defined.

III. The SDW-MWF

The MSE between the output of a BF, $\mathbf{w}'$, and the desired signal is $\mathrm{E}\{|s_d(\ell) - (\mathbf{w}')^\dagger \mathbf{z}(\ell)|^2\}$. The SDW-MWF BF is designed to minimize a weighted version of the MSE and its goal is to enhance the desired signal $s_d(\ell)$. It is defined as the solution of the following minimization problem:

$$\begin{aligned}\mathbf{w} &\triangleq \underset{\mathbf{w}'}{\operatorname{argmin}} |1 - ((\mathbf{w}')^\dagger \mathbf{h}_d)|^2 \delta_d + \mu (\mathbf{w}')^\dagger \mathbf{\Phi}_{vv} \mathbf{w}' \\ &= \frac{\mathbf{\Phi}_{vv}^{-1} \mathbf{h}_d}{\mathbf{h}_d^\dagger \mathbf{\Phi}_{vv}^{-1} \mathbf{h}_d + \frac{\mu}{\delta_d}}\end{aligned} \tag{4}$$

where μ is a non-negative parameter which controls the tradeoff between the interference reduction and the desired signal distortion. For $\mu = 1$ the classical Wiener filter [7] (MMSE) is obtained. At the limit $\mu \to 0$ the MVDR-BF is reached, and no distortion is introduced to the desired signal.

Next, we define two criteria of BFs. The SIR at the output of a BF, $\mathbf{w}$, is denoted κ and is defined as the ratio of the powers of the desired signal and the interference signals at the beamformer output, i.e.:

$$\kappa \triangleq \frac{\delta_d |\mathbf{w}^\dagger \mathbf{h}_d|^2}{\mathbf{w}^\dagger \mathbf{\Phi}_{vv} \mathbf{w}}. \tag{5}$$

The white noise gain [24] is denoted ξ and is defined as the SNR gain of the BF for a spatially white noise. It equals:

$$\xi \triangleq \frac{|\mathbf{w}^\dagger \mathbf{h}_d|^2}{\|\mathbf{w}\|^2}. \tag{6}$$

By substituting (4) in (5) and (6), we obtain expressions for the performance criteria of the SDW-MWF:

$$\kappa = \delta_d \mathbf{h}_d^\dagger \mathbf{\Phi}_{vv}^{-1} \mathbf{h}_d \tag{7a}$$

$$\xi = \frac{\left(\mathbf{h}_d^\dagger \mathbf{\Phi}_{vv}^{-1} \mathbf{h}_d\right)^2}{\mathbf{h}_d^\dagger \mathbf{\Phi}_{vv}^{-2} \mathbf{h}_d}. \tag{7b}$$

Note, that both expressions do not depend on μ. This can be attributed to the fact that SDW-MWF equals the MVDR-BF followed by a single channel SDW-MWF [25], [8] with the parameter μ. Hence, the SIR and white noise gain at the output of the SDW-MWF equal to their respective quantities at the output of an MVDR-BF (locally, per frequency bin).

In the next section, a statistical model for the ATF is presented. From this model we will derive the statistics of the SIR and the white noise gain criteria for various noise fields.

IV. ATF Statistics

In the following sections approximations for the first and second moments of the ATF are derived. We will show that, under several assumptions, the ATFs relating a source with the microphone array can be modeled as independent identically distributed (i.i.d.) complex Gaussian random variables (RVs) with zero mean and a variance which depends on the properties of the enclosure. Furthermore, ATFs of different sources are shown to be uncorrelated.

A. Single ATF Statistics

Let h be an ATF relating a coherent point source signal, located at $\mathbf{r}$, and the mth microphone, located at $\mathbf{r}^m$. The ATF is comprised of two components: the direct arrival ATF and the reverberant component ATF. Since the microphone location, $\mathbf{r}^m$, is random, the ATF is also a complex RV. Explicitly:

$$h \triangleq \bar{h} + \hat{h} \tag{8}$$

where $\bar{h}$ and $\hat{h}$ denote direct arrival and reverberant components, respectively. We assume that the direct arrival and the reverberant ATFs are uncorrelated. Define the room volume and surface area as:

$$V \triangleq D_x \times D_y \times D_z \tag{9a}$$

$$A \triangleq 2(D_x \times D_y + D_x \times D_z + D_y \times D_z) \tag{9b}$$

and denote the reverberation time as T_{60}. Adopting the ATF model of Schroeder [19], [22], $\hat{h}$ is modeled as a complex Gaussian RV:

$$\hat{h} \sim \mathcal{CN}(0, \hat{\alpha}) \tag{10}$$

where

$$\hat{\alpha} \triangleq \frac{1 - \varepsilon}{\pi \varepsilon A} \tag{11}$$

and

$$\varepsilon \triangleq \frac{0.161 V}{A T_{60}} \tag{12}$$

is the exponential decay rate of the RIR tail. The latter model is valid under the following assumptions: first, the signal wavelength is much smaller than the room dimensions; second, microphones and sources are at least half a wavelength away from the walls; third, the signal frequency is above the Schroeder frequency, defined as:

$$f_{\text{Schroeder}} \triangleq 2000 \sqrt{\frac{T_{60}}{V}}. \tag{13}$$

In typical acoustic scenarios $f_{\text{Schroeder}}$ is in the order of a few hundred Hz.

The direct arrival ATF is given in a polar representation by

$$\bar{h} \triangleq \bar{a} \exp(\mathrm{j}\bar{\phi}) \tag{14}$$

where $\bar{a}$ and $\exp(\mathrm{j}\bar{\phi})$ are the amplitude and phase responses, respectively. Assuming spherical wave propagation:

$$\bar{a} = \begin{cases} 1; & \|\mathbf{r}\| \leq \frac{1}{4\pi} \\ \frac{1}{4\pi\|\mathbf{r}-\mathbf{r}^m\|}; & \frac{1}{4\pi} < \|\mathbf{r}\| \end{cases} \tag{15a}$$

$$\bar{\phi} = \frac{2\pi\|\mathbf{r}-\mathbf{r}^m\|}{\lambda_k} \tag{15b}$$

where $\lambda_k = (cK)/(kf_s)$ is the wavelength corresponding to the kth frequency bin, K is the STFT window length, f_s is the sampling frequency rate and c is the sound velocity. Furthermore, since (15a) is not physically meaningful for $\|\mathbf{r}-\mathbf{r}^m\| \to 0$, we limit the amplitude response to $\bar{a} = 1$ for $\|\mathbf{r}-\mathbf{r}^m\| < (1)/(4\pi)$. Without loss of generality, consider that the source is located at the origin $\mathbf{r} = \mathbf{0}$, and that a sphere with radius $\bar{r}$, centered at the origin, is within the room volume. The sphere radius is chosen such that $\bar{r} \gg \lambda_k$. Since multiple 2π phase cycles are repeated while propagating in the sphere, the amplitude and phase responses, $\bar{a}$ and $\exp(\mathrm{j}\bar{\phi})$ can be approximated as uncorrelated inside the sphere. We verify this approximation in Section IV-C. Note, that the direct arrival component is a stochastic variable, since it is a function of the microphone location which is random, and that the reverberant component is stochastic under Schroeder's model. The subsequent expectation operations should be interpreted accordingly.

Moreover, the mean phase response is approximately

$$\mathrm{E}\{\exp(\mathrm{j}\bar{\phi})\} \approx 0 \tag{16}$$

and the mean direct arrival ATF is:

$$\begin{aligned}\mathrm{E}\{\bar{h}\} &= \mathrm{E}\{\bar{a}\}\mathrm{E}\{\exp(\mathrm{j}\bar{\phi})\} \\ &= 0.\end{aligned} \tag{17}$$

These results are also verified in Section IV-C.

Considering a microphone which location is uniformly distributed in a sphere with radius $\bar{r}$ and using (15a), the variance of the direct arrival equals the solution of the following integral in spherical coordinates:

$$\begin{aligned}\bar{\alpha} &= \mathrm{E}\{\bar{a}^2 | \|\mathbf{r}^m\| < \bar{r}\} \\ &= \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\int_{r=0}^{\bar{r}} \frac{3}{4\pi\bar{r}^3}\bar{a}^2 r^2 \sin(\theta)\mathrm{d}r\mathrm{d}\theta\mathrm{d}\phi \\ &= \frac{6\pi\bar{r}-1}{32\pi^3\bar{r}^3}.\end{aligned} \tag{18}$$

Combining (10) and (17), the mean ATF equals:

$$\mathrm{E}\{h\} = 0. \tag{19}$$

Denote the variance of the ATF:

$$\alpha \triangleq \mathrm{E}\{|h|^2\}. \tag{20}$$

Using the law of total probability, (20) can be written as:

$$\begin{aligned}\alpha = \Pr(\|\mathbf{r}^m\| < \bar{r})\mathrm{E}\{|h|^2 | \|\mathbf{r}^m\| < \bar{r}\} \\ + \Pr(\|\mathbf{r}^m\| \geq \bar{r})\mathrm{E}\{|h|^2 | \bar{r} \leq \|\mathbf{r}^m\|\}.\end{aligned} \tag{21}$$

Denote the critical distance, the distance from the source at which the powers of the direct arrival and the reverberant components are equal, as r_c. Kuttruff [22] derived an expression for the critical distance:

$$r_c \triangleq \sqrt{\frac{V}{100\pi T_{60}}}. \tag{22}$$

We assume that $\bar{r} \gg r_c$, i.e., the sphere radius $\bar{r}$ is much larger than the critical distance. Hence $|\bar{h}| \ll |\hat{h}|$ (the direct arrival ATF is negligible compared to the reverberant ATF component) for $\|\mathbf{r}^m\| > \bar{r}$, and (21) is approximately:

$$\begin{aligned}\alpha &\approx \Pr(\|\mathbf{r}^m\| < \bar{r})\mathrm{E}\{|\bar{h}|^2 + |\hat{h}|^2 | \|\mathbf{r}^m\| < \bar{r}\} \\ &\quad + \Pr(\|\mathbf{r}^m\| \geq \bar{r})\mathrm{E}\{|\hat{h}|^2 | \bar{r} \leq \|\mathbf{r}^m\|\} \\ &= \Pr(\|\mathbf{r}^m\| < \bar{r})\mathrm{E}\{|\bar{h}|^2 | \|\mathbf{r}^m\| < \bar{r}\} + \mathrm{E}\{|\hat{h}|^2\}.\end{aligned} \tag{23}$$

Note, that we utilized the fact that $\bar{h}$ and $\hat{h}$ are uncorrelated. Substituting (18) and (10) in (23) and noting that $\Pr(\|\mathbf{r}^m\| < \bar{r}) = (4\pi\bar{r}^3)/(3V)$ yields:

$$\alpha = \frac{4\pi\bar{r}^3}{3V}\bar{\alpha} + \hat{\alpha}. \tag{24}$$

We return to the original coordinate system, centered in the room, and consider the statistics of an ATF vector $\mathbf{h}$. Since the microphone locations are i.i.d., and since h^m, the mth element in $\mathbf{h}$, depends on the location of the mth microphone, $\mathbf{r}^m$, we conclude that h^m; $m = 1, \ldots, M$ are i.i.d..

B. Cross -Covariance of ATFs

In this section we model the cross-covariance of the ATFs relating two sources located at $\mathbf{r}_1$ and $\mathbf{r}_2$ with a microphone randomly located at $\mathbf{r}^m$. The covariance is comprised of the sum of the direct arrival and reverberant component covariances:

$$\mathrm{E}\{h_1 h_2^*\} = \mathrm{E}\{\bar{h}_1\bar{h}_2^*\} + \mathrm{E}\{\hat{h}_1\hat{h}_2^*\} \tag{25}$$

where h_i, $\bar{h}_i$ and $\hat{h}_i$ are, respectively, the total ATF, the direct arrival ATF and the reverberant ATF for sources $i = 1, 2$. Similarly to (14), the amplitude and phase components of the ith source direct arrival ATF are defined as $\bar{h}_i = \bar{a}_i \exp(\mathrm{j}\bar{\phi}_i)$.

First, let us examine the covariance of the reverberant ATFs. Schroeder models the reverberant ATF as the sum of multiple independent reflections arriving from all directions. Hence, their coherence, defined as $(\mathrm{E}\{\hat{h}_1\hat{h}_2^*\})/(\hat{\alpha})$, is equivalent to the coherence between two microphones located at $\mathbf{r}_1$ and $\mathbf{r}_2$ in a diffuse sound field (comprised of multiple uncorrelated sources radiating from a surrounding sphere) [26]. Explicitly, the covariance equals:

$$\mathrm{E}\{\hat{h}_1\hat{h}_2^*\} = \hat{\alpha}\,\mathrm{sinc}\left(\frac{2\pi\|\mathbf{r}_1 - \mathbf{r}_2\|}{\lambda_k}\right) \tag{26}$$

where $\mathrm{sinc}(x) = (\sin(x))/(x)$. Assuming that the distance between the sources is much larger than the wavelength, i.e., $\|\mathbf{r}_1 - \mathbf{r}_2\| \gg \lambda_k$, the covariance between the reverberant component is approximately:

$$\mathrm{E}\{\hat{h}_1\hat{h}_2^*\} \approx 0. \tag{27}$$

Consider the expectation of the inner product of direct arrival ATFs in (25), i.e., $\mathrm{E}\{\bar{h}_1\bar{h}_2^*\} = \mathrm{E}\{\bar{a}_1\bar{a}_2\exp(\mathrm{j}(\bar{\phi}_1 - \bar{\phi}_2))\}$. Again, since $\|\mathbf{r}_1 - \mathbf{r}_2\| \gg \lambda_k$ and by applying same considerations as leading to (17), we approximate that the phase and amplitude in the last expression are uncorrelated, and that $\mathrm{E}\{\exp(\mathrm{j}(\bar{\phi}_1 - \bar{\phi}_2))\} \approx 0$. Hence:

$$\mathrm{E}\{\bar{h}_1\bar{h}_2^*\} \approx 0. \tag{28}$$

Finally, substituting (27) and (28) in (25), we conclude that the ATFs are uncorrelated:

$$\mathrm{E}\{h_1h_2^*\} = 0. \tag{29}$$

C. Model Verification

In order to verify the proposed simplified model we have conducted several different Monte-Carlo experiments. First, the model of the ATF statistics is examined for various reverberation times and enclosure dimensions. The theoretical model was calculated with the parameter $\bar{r} = 2r_c$. In the first experiment, the location of a single microphone was uniformly randomized in a 4 m × 4 m × 3 m room with a reverberation time in the range $[0.2\ \mathrm{sec}, 0.3\ \mathrm{sec}, \ldots, 0.8\ \mathrm{sec}]$. In the second experiment, we examined the relation between the ATF statistics and the room dimensions. We set the reverberation time to 0.3 sec and examined different room sizes: $(2 + 0.5i)\ \mathrm{m} \times (2 + 0.5i)\ \mathrm{m} \times (2.2 + 0.1i)\ \mathrm{m}$ for $i = 0, \ldots, 8$. For each room configuration in both experiments, the locations of a single source and a microphone were uniformly randomized in the room. The locations of the single source were randomized 4 times, and for each case 100 locations of the microphone were randomized. For each case, the direct arrival and the tail of the ATF as well as the complete ATF were generated using [27]. The normalized error of the empirical mean of the ATF, is defined as $(|\langle h\rangle - \mathrm{E}\{h\}|^2)/(\mathrm{E}\{|h|^2\}) = (|\langle h\rangle|^2)/(\mathrm{E}\{|h|^2\})$, where $\langle\bullet\rangle$ denotes the empirical average. As we expect that $\mathrm{E}\{h\} \approx 0$, as in (19), the ratio $(|\langle h\rangle|^2)/(\mathrm{E}\{|h|^2\})$ is considered to verify this approximation in both experiments. For all tested reverberation times and room dimensions the normalized error, averaged over all considered scenarios, is -20 dB, and clearly the approximation $\mathrm{E}\{h\} \approx 0$ holds. The theoretical and empirical (denoted emp.) variances of direct arrival, reverberant (denoted rev.) and total ATFs, i.e., $\bar{\alpha}$, $\hat{\alpha}$ and α, are depicted in Figs. 1, 2 for both experiments, respectively. The empirical variances were averaged over the frequency range of [300 Hz, 3700 Hz]. Note that the reverberation time, T_{60}, affects the variance of the direct arrival ATF, (18), from the setting $\bar{r} = 2r_c$ and the definition of the critical distance (22). In these figures, it is clearly depicted that the model for $\hat{\alpha}$, the variance of the reverberant component, is accurate, whereas the model for $\bar{\alpha}$, the variance of the direct arrival, demonstrates small mismatch. The model for α, the variance of the total ATF, is accurate, since it is mostly affected by the reverberant component. These results also apply when considering a specific frequency (in the specified range), instead of averaging over all frequencies.

In the third experiment, we verified our theoretical result stating that the ATFs are uncorrelated. The room dimensions were set to 3 m × 3 m × 2.4 m. We tested the statement for various reverberation times 0.2 sec, 0.4 sec and 0.6 sec, and for different distances between the sources $r_{1,2} \triangleq \|\mathbf{r}_1 - \mathbf{r}_2\| = 0.2$ m, 0.6 m, …, 1.8 m. For each room configuration 4 locations of the sources were randomly selected, and for each source

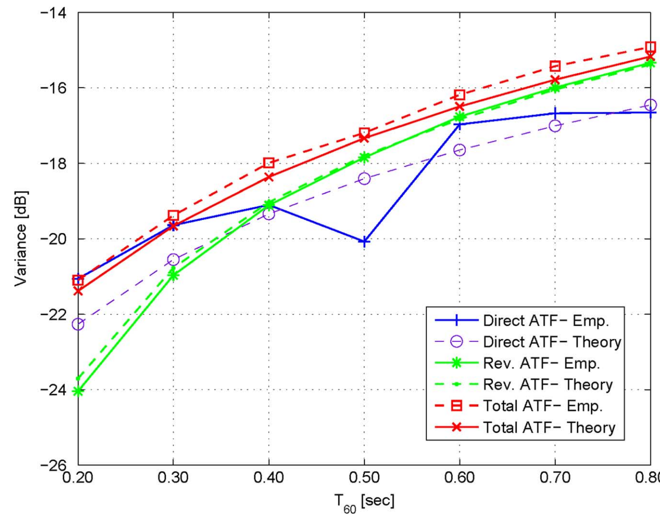


Fig. 1. Empirical and theoretical ATF variances versus reverberation time.

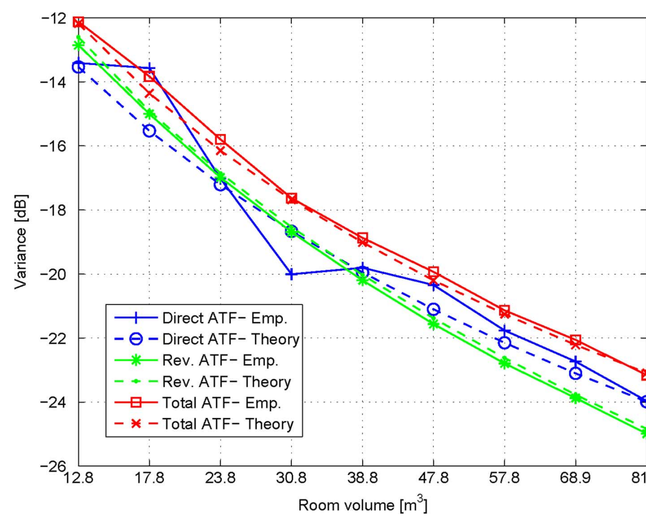


Fig. 2. Empirical and theoretical ATF variances versus room dimensions.

location 100 different locations of microphones were randomized. The empirical coherence of the total ATFs is defined as $\mathrm{coh}(h_1, h_2) \triangleq (\langle h_1h_2^*\rangle)/(\sqrt{\langle|h_1|^2\rangle\langle|h_2|^2\rangle})$, where we note that h_1 and h_2 are zero mean. In a similar manner, we define $\mathrm{coh}(\bar{h}_1, \bar{h}_2)$, $\mathrm{coh}(\hat{h}_1, \hat{h}_2)$. The empirical coherence of the direct ATFs, the reverberant ATFs and the total ATFs for all tested reverberation times and distances between sources are lower than -30 dB (averaged over all frequencies). The empirical coherence versus frequency in the case of $T_{60} = 0.4$ sec and $r_{1,2} = 0.2$ m is depicted in Fig. 3. The results of this evidently verify the assumption that the ATFs of different sources are uncorrelated. In all experiments a sampling rate of 8 kHz is used.

V. Beamformers Performance

In this section, we analyze the performance of the SDW-MWF in various noise fields. We derive *reliability* measures for the SIR and white noise gain criteria. The reliability of an SIR level of κ_0 is defined as the probability that the output SIR will exceed κ_0, explicitly:

$$R_\kappa(\kappa_0) \triangleq \Pr(\kappa \geq \kappa_0). \tag{30}$$

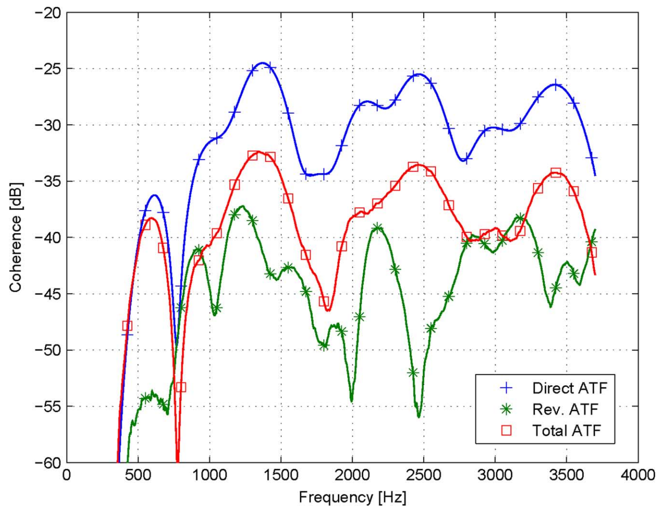


Fig. 3. Empirical coherence between ATFs versus frequency.

Similarly, the reliability of a white noise gain level of ξ_0, is defined as the probability that the white noise gain will exceed ξ_0:

$$R_\xi(\xi_0) \triangleq \Pr(\xi \geq \xi_0). \tag{31}$$

The reliability functions can be used to predict the performance of the BF in the WASN. Moreover, they can be used to determine the number of microphones that should be used in order to meet a predefined performance level. However, as these measures are statistical, for any microphones location realization, a non-zero probability that the desired performance level will not be met exists.

A. Coherent Interference Signals $P < M$

Let $s_{i,1}(\ell), \ldots, s_{i,P}(\ell)$ be $P < M$ coherent noise sources located at $\mathbf{r}_{i,1}, \ldots, \mathbf{r}_{i,P}$, respectively. Denote the covariance matrix of the interference signals as:

$$\boldsymbol{\Delta}_i \triangleq \operatorname{diag}\{\delta_{i,1}, \ldots, \delta_{i,P}\} \tag{32}$$

where $\delta_{i,p} = \mathrm{E}\{|s_{i,p}(\ell)|^2\}$ is the variance of the pth source, for $p = 1, \ldots, P$, and let

$$\mathbf{h}_{i,p} \triangleq \begin{bmatrix} h_{i,p}^1 & \cdots & h_{i,p}^M \end{bmatrix}^T \tag{33}$$

be the ATF relating the pth interfering source and the microphones. The received interference signals vector, in the case of coherent interference signals, is therefore given by:

$$\mathbf{v}(\ell) = \mathbf{H}_i \mathbf{s}_i(\ell) + \mathbf{u}(\ell) \tag{34}$$

where

$$\mathbf{H}_i = \begin{bmatrix} \mathbf{h}_{i,1} & \cdots & \mathbf{h}_{i,P} \end{bmatrix} \tag{35a}$$

$$\mathbf{s}_i(\ell) = \begin{bmatrix} s_{i,1}(\ell) & \cdots & s_{i,P}(\ell) \end{bmatrix}^T \tag{35b}$$

and $\mathbf{u}(\ell)$ is a complex Gaussian sensors noise with covariance $\delta_u \mathbf{I}$, i.e., $\mathbf{u}(\ell) \sim \mathcal{CN}(\mathbf{0}, \delta_u \mathbf{I})$. Therefore, the covariance matrix of the received interference signals is:

$$\boldsymbol{\Phi}_{vv} = \mathbf{H}_i \boldsymbol{\Delta}_i \mathbf{H}_i^\dagger + \delta_u \mathbf{I}. \tag{36}$$

Consider the expression $\mathbf{h}_d^\dagger \boldsymbol{\Phi}_{vv}^{-1} \mathbf{h}_d$ which appears in the SIR and white noise gain criteria of the SDW-MWF in (7a) and (7b). Applying the Woodbury identity to $\boldsymbol{\Phi}_{vv}^{-1}$ as defined in (36) yields:

$$\boldsymbol{\Phi}_{vv}^{-1} = \delta_u^{-1} \mathbf{I}_M - \delta_u^{-1} \mathbf{H}_i \left(\mathbf{I}_P + \delta_u \left(\mathbf{H}_i^\dagger \mathbf{H}_i \right)^{-1} \boldsymbol{\Delta}_i^{-1} \right)^{-1} \times \left(\mathbf{H}_i^\dagger \mathbf{H}_i \right)^{-1} \mathbf{H}_i^\dagger. \tag{37}$$

Now, assuming that the power of the coherent interference signals is much larger than the variance of the sensors noise, and further assuming that $\mathbf{I}_P + \delta_u (\mathbf{H}_i^\dagger \mathbf{H}_i)^{-1} \boldsymbol{\Delta}_i^{-1} \approx \mathbf{I}_P$, (37) can be approximated as:

$$\boldsymbol{\Phi}_{vv}^{-1} \approx \delta_u^{-1} \left(\mathbf{I}_M - \mathbf{H}_i \left(\mathbf{H}_i^\dagger \mathbf{H}_i \right)^{-1} \mathbf{H}_i^\dagger \right). \tag{38}$$

Note that, $\boldsymbol{\Phi}_{vv}^{-1}$ is approximately a projection matrix to the null-subspace of $\mathbf{H}_i$, scaled by δ_u^{-1}.

Let

$$\mathbf{H}_i = \boldsymbol{\Psi} \boldsymbol{\Omega} \boldsymbol{\Theta}^\dagger \tag{39}$$

be the singular value decomposition (SVD) of $\mathbf{H}_i$. Substituting (39) in (38), we obtain the more compact expression:

$$\boldsymbol{\Phi}_{vv}^{-1} = \delta_u^{-1} \dot{\boldsymbol{\Psi}} \dot{\boldsymbol{\Psi}}^\dagger \tag{40}$$

where $\dot{\boldsymbol{\Psi}}$ is an $M \times (M-P)$ matrix comprising the $M-P$ last columns of $\boldsymbol{\Psi}$, associated with the zero singular values, which span the null-subspace of $\mathbf{H}_i$.

Defining

$$\boldsymbol{\rho}_c \triangleq \dot{\boldsymbol{\Psi}}^\dagger \mathbf{h}_d \tag{41}$$

and substituting (40) and (41) in (7a) and (7b) yields the simplified criteria expressions:

$$\kappa_c = \frac{\delta_d}{\delta_u} \|\boldsymbol{\rho}_c\|^2 \tag{42a}$$

$$\xi_c = \|\boldsymbol{\rho}_c\|^2 \tag{42b}$$

where we denote the SIR and white noise gain criteria in the coherent noise field (for $P < M$) as κ_c and ξ_c, respectively. Note that both criteria depend on $\|\boldsymbol{\rho}_c\|^2$ with different coefficient multipliers. Now, we turn to analyze their statistics.

Denote the nth column of $\dot{\boldsymbol{\Psi}}$ and the nth element of $\boldsymbol{\rho}_c$ as $\dot{\boldsymbol{\psi}}_n$ and $\rho_{c,n}$, respectively, for $n = 1, \ldots, (M-P)$. A single element $\rho_{c,n}$ is obtained by

$$\rho_{c,n} = \dot{\boldsymbol{\psi}}_n^\dagger \mathbf{h}_d \tag{43}$$

which is a linear combination of the uncorrelated elements of $\mathbf{h}_d$. From the unitarity of $\boldsymbol{\Psi}$, we conclude that $\boldsymbol{\rho}_c$ is a vector of uncorrelated variables:

$$\begin{aligned} \mathrm{E}\{\boldsymbol{\rho}_c \boldsymbol{\rho}_c^\dagger\} &= \mathrm{E}\{\dot{\boldsymbol{\Psi}}_i^\dagger \mathbf{h}_d \mathbf{h}_d^\dagger \dot{\boldsymbol{\Psi}}_i\} \\ &= \alpha \mathbf{I}_{M-P}. \end{aligned} \tag{44}$$

Now, since $(M-P) \gg 1$ the central limit theorem (CLT) conditions hold and hence we argue that the distribution of the random variable $\rho_{c,n}$; $n = 1, \ldots, M-P$ converges to the complex normal distribution $\rho_{c,n} \sim \mathcal{CN}(0, \alpha)$, where α is defined in (24):

$$\boldsymbol{\rho}_c \sim \mathcal{CN}(\mathbf{0}, \alpha \mathbf{I}_{M-P}). \quad (45)$$

It is easily concluded that the elements of $\boldsymbol{\rho}_c$ are i.i.d.

Define

$$\eta_c \triangleq \frac{2}{\alpha} \|\boldsymbol{\rho}_c\|^2. \quad (46)$$

It is a Chi-square RV with $2(M-P)$ degrees of freedom, i.e.,

$$\eta_c \sim \chi^2(2(M-P)). \quad (47)$$

From (46) we have:

$$\|\boldsymbol{\rho}_c\|^2 = \frac{\alpha}{2} \eta_c. \quad (48)$$

Substituting (48) in (42a) and (42b) yields alternative expressions for the performance criteria:

$$\kappa_c = \frac{\delta_d}{\delta_u} \frac{\alpha}{2} \eta_c \quad (49a)$$

$$\xi_c = \frac{\alpha}{2} \eta_c. \quad (49b)$$

Using the probability distribution function (p.d.f.) of η_c, the average SIR and white noise gain, denoted $\bar{\kappa}_c$ and $\bar{\xi}_c$, are given by:

$$\bar{\kappa}_c \triangleq \mathrm{E}\{\kappa_c\} = \frac{\alpha \delta_d}{\delta_u} (M-P) \quad (50a)$$

$$\bar{\xi}_c \triangleq \mathrm{E}\{\xi_c\} = \alpha (M-P) \quad (50b)$$

where we substituted $\mathrm{E}\{\eta_c\} = 2(M-P)$. Note, that the latter averages of the SIR and white noise gain are linear with the number of microphones.

From (49a) it is clear that κ_c is a scaled version of a Chi-square RV with $2(M-P)$ degrees of freedom. Hence, its reliability (30) can be calculated as:

$$R_{\kappa,c}(\kappa_0) = 1 - F_{\eta,c}\left(\frac{2}{\alpha} \frac{\delta_u}{\delta_d} \kappa_0\right) \quad (51)$$

where

$$F_{\eta,c}(\eta_0) \triangleq \Pr(\eta_c \leq \eta_0) = \frac{\gamma_f\left(M-P, \frac{\eta_0}{2}\right)}{\Gamma_f(M-P)} \quad (52)$$

is the cumulative distribution function of a Chi-square RV with $2(M-P)$ degrees of freedom, Γ_f is the Gamma function and γ_f is the lower incomplete Gamma function.

Similarly, the reliability of the white noise gain (49b) is:

$$R_{\xi,c}(\xi_0) = 1 - F_{\eta,c}\left(\frac{2}{\alpha} \xi_0\right). \quad (53)$$

B. *Diffuse Sound Field*

In this section we derive the performance of an SDW-MWF in a diffuse sound field. This noise field is can be modeled by numerous statistically independent noise sources arriving from all directions simultaneously ($P \gg M$). It is a common noise field in reverberant environments, cocktail party and car scenarios [28]. The covariance $\Phi_{vv}(m, m')$ of a diffused noise between the noise components received at the mth and the m'th microphones equals:

$$\Phi_{vv}(m, m') = \delta_{\mathrm{dif}} \mathrm{sinc}\left(\frac{2\pi \|\mathbf{r}^m - \mathbf{r}^{m'}\|}{\lambda_k}\right) \quad (54)$$

where δ_{dif} denotes the variance of the diffuse sound field received at each microphone. Note that the covariance is a RV due to the random microphone locations.

Consider, the average covariance $\mathrm{E}\{\Phi_{vv}(m, m')\}$ in a sphere with radius $R \gg \lambda_k$ around $\mathbf{r}^{m'}$. Assuming that the mth microphone is randomly located in the sphere with a uniform distribution, the expectation can be formulated as:

$$\mathrm{E}\{\Phi_{vv}(m, m')\} = \iiint \frac{\delta_{\mathrm{dif}} \,\mathrm{sinc}\left(\frac{2\pi r}{\lambda_k}\right) r \sin(\theta) \mathrm{d}r \mathrm{d}\theta \mathrm{d}\phi}{4/3 \pi R^3} = \delta_{\mathrm{dif}} \times \begin{cases} 1; & m = m' \\ \frac{3 \sin^2\left(\frac{\pi R}{\lambda_k}\right)}{2R}; & m \neq m' \end{cases}. \quad (55)$$

Now, since $R \gg \lambda_k$ we can approximate:

$$\mathrm{E}\{\Phi_{vv}(m, m')\} \approx \delta_{\mathrm{dif}} \mathbf{I}. \quad (56)$$

Since the enclosure is much larger than λ_k, we assume that on average the distance between any pair of microphones is larger than λ_k, and propose the approximation:

$$\boldsymbol{\Phi}_{vv} \approx \delta_{\mathrm{dif}} \mathbf{I}. \quad (57)$$

Define

$$\boldsymbol{\rho}_{\mathrm{dif}} \triangleq \sqrt{\delta_{\mathrm{dif}}} \left(\boldsymbol{\Phi}_{vv}^{-1/2}\right)^{\dagger} \mathbf{h}_d \quad (58)$$

where $\boldsymbol{\Phi}_{vv}^{-1} = \boldsymbol{\Phi}_{vv}^{-1/2} (\boldsymbol{\Phi}_{vv}^{-1/2})^{\dagger}$ is the Cholesky decomposition. Since, in most cases the power of the reverberant component dominates the ATF, we propose to model $\mathbf{h}_d \sim \mathcal{CN}(\mathbf{0}, \alpha \mathbf{I}_{M \times M})$, and by using (57) to model $\boldsymbol{\rho}_{\mathrm{dif}}$ as an $M \times 1$ complex Gaussian RV with the probability distribution:

$$\boldsymbol{\rho}_{\mathrm{dif}} \sim \mathcal{CN}(\mathbf{0}, \alpha \mathbf{I}_{M \times M}). \quad (59)$$

Define

$$\eta_{\mathrm{dif}} \triangleq \frac{2}{\alpha} \|\boldsymbol{\rho}_{\mathrm{dif}}\|^2 \quad (60)$$

and note that η_{dif} is a Chi-square RV with $2M$ degrees of freedom, i.e.,

$$\eta_{\mathrm{dif}} \sim \chi^2(2M). \quad (61)$$

Substituting (57), (58) and (60) in (7a) and (7b) yields:

$$\kappa_{\mathrm{dif}} = \frac{\delta_d}{\delta_{\mathrm{dif}}} \frac{\alpha}{2} \eta_{\mathrm{dif}} \quad (62a)$$

$$\xi_{\mathrm{dif}} = \frac{\alpha}{2} \eta_{\mathrm{dif}} \quad (62b)$$

where we have applied the following approximation resulting from (57):

$$\mathbf{\Phi}_{vv}^{-2} \approx \frac{1}{\delta_{\text{dif}}} \mathbf{\Phi}_{vv}^{-1}. \tag{63}$$

The average SIR and white noise gain in the diffuse sound field case are given by:

$$\bar{\kappa}_{\text{dif}} \triangleq \mathrm{E}\{\kappa_{\text{dif}}\} = \frac{\alpha \delta_d}{\delta_{\text{dif}}} M \tag{64a}$$

$$\bar{\xi}_{\text{dif}} \triangleq \mathrm{E}\{\xi_{\text{dif}}\} = \alpha M \tag{64b}$$

where we substitute $\mathrm{E}\{\eta_{\text{dif}}\} = 2M$. Note, that as in the coherent interference signals case, the latter averages of the SIR and white noise gain are linear with the number of microphones. Van Trees [29] showed that, for spatially white noise, the SIR linearly increases with the number of microphones in the deterministic case (when the microphone locations are not random). Since we show that the diffuse sound field covariance matrix can be approximated by a scaled identity matrix, we obtain a similar result for randomly located microphones.

Similarly to (51), (53) the reliability of κ_{dif} and ξ_{dif} are given by:

$$R_{\kappa,\text{dif}}(\kappa_0) = 1 - F_{\eta,\text{dif}}\left(\frac{2}{\alpha} \frac{\delta_{\text{dif}}}{\delta_d} \kappa_0\right) \tag{65a}$$

$$R_{\xi,\text{dif}}(\xi_0) = 1 - F_{\eta,\text{dif}}\left(\frac{2}{\alpha} \xi_0\right) \tag{65b}$$

where

$$F_{\eta,\text{dif}}(\eta_0) \triangleq \Pr(\eta_{\text{dif}} \leq \eta_0) = \frac{\gamma_f\left(M, \frac{\eta_0}{2}\right)}{\Gamma_f(M)} \tag{66}$$

is the cumulative distribution function of the $2M$ degrees of freedom Chi-square RV η_{dif}.

VI. BF Model Verification

A. Coherent Interference Signals $P < M$

We carry out experiments to verify the theoretical model for the statistics of κ_c and ξ_c in the case of coherent noise sources. The room dimensions are set to 4 m × 4 m × 3 m, the number of microphones is $M = 5$, the sampling rate is 8 kHz, and the discrete Fourier transform (DFT) size is $K = 8192$. The number of interference signals is in the range of $P = 1, 2, 3, 4$ and the reverberation time can take the values 0.2 sec, 0.4 sec, 0.6 sec, 0.8 sec. The received microphone signals comprise of a coherent desired source, modeled as a 6th order auto regressive (AR) random process, coherent noise sources, modeled as an AR(1) random processes, and sensors noise. We simulate the spectra of the signals, and substitute them in the derived formulas. The average SNR of the desired source, and the average INR of each of the coherent noise sources are set to 90 dB at the microphones. The locations of the desired source and the interference signals are randomly selected in 4 scenarios. In each scenario 100 microphones positions are drawn with a 3D uniform

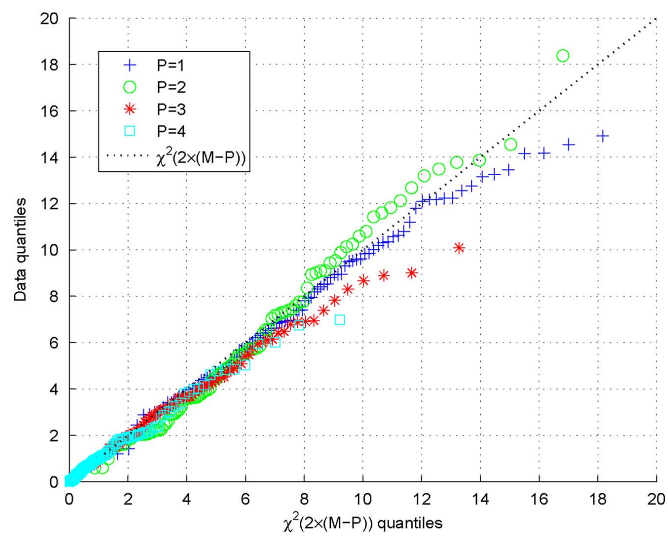


Fig. 4. Quantile-quantile probability plot of $(2\delta_u)/(\alpha\delta_d)\kappa_c$ versus $\chi^2(2(M-P))$ distribution with various numbers of coherent interference signals, $P = 1, \ldots, 4$, and for a reverberation time of $T_{60} = 0.4$ sec.

distribution. In each Monte-Carlo experiment, and per each frequency, the SDW-MWF with $\mu = 1$ is calculated, and its SIR and white noise gain are recorded. The normalized errors of the average SIR and white noise gain, i.e., κ_c and ξ_c, defined as $((\langle\kappa_c\rangle - \bar{\kappa}_c)^2)/(\bar{\kappa}_c^2)$ and $((\langle\xi_c\rangle - \bar{\xi}_c)^2)/(\bar{\xi}_c^2)$, respectively, are -20 dB for low and medium reverberation times, 0.2 sec, 0.4 sec, and for all numbers of interference signals scenarios. For higher reverberation times, 0.6 sec, 0.8 sec, the measured normalized errors are a bit higher, -15 dB. Evidently, the formulas for the average criteria (50a), (50b) are valid.

The following figures correspond to one of the desired source and interference signals constellations at frequency 2 kHz. Similar results are obtained for other scenarios and frequencies. In the derivation of the theoretical model, we argued that κ_c is a scaled $\chi^2(2(M-P))$ RV. The quantile-quantile probability plots of $(2\delta_u)/(\alpha\delta_d)\kappa_c$ versus the $\chi^2(2(M-P))$ distribution is depicted in Fig. 4 for reverberation time $T_{60} = 0.4$ sec, and for various numbers of interference signals $P = 1, \ldots, 4$. From this figure, the Chi-square distribution with $2(M-P)$ degrees of freedom of the scaled κ_c can be verified. We also verify that $\boldsymbol{\rho}_c$ is an $(M-P) \times 1$ complex normal vector. The reliability function of the SIR, i.e., $R_{\kappa,c}$, versus the SIR improvement (defined as the ratio of the output and input SIR) for $T_{60} = 0.4$ sec is depicted in Fig. 5. Clearly from this figure, the reliability function of the SIR is verified. As expected, the reliability of the white noise gain demonstrates similar behavior.

The reliability functions of the SIR and white noise gain were measured for all combinations of $T_{60} = 0.2, 0.4, 0.6, 0.8$ sec and $P = 1, 2, 3, 4$ interference signals for various values of $\mu = 1, 10, 100$. Correspondingly to derivation in (51), (53), the measured reliability criteria are independent of the parameter μ.

Now, we wish to verify the effect of the number of microphones M on the reliability measures and of the SIR and white noise gain in the coherent interference signals case. We use the same room dimensions as above, and set the reverberation time to $T_{60} = 0.4$ sec. We test 4 different constellations of a desired source and a single interference signal. For each constellation

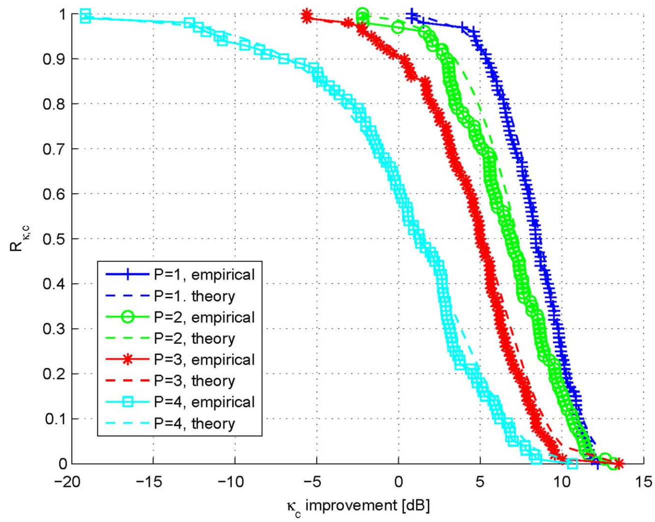


Fig. 5. Reliability function of SIR, $R_{\kappa,c}$, versus SIR improvement with various numbers of coherent interference signals, $P = 1, \ldots, 4$, and for a reverberation time of $T_{60} = 0.4$ sec.

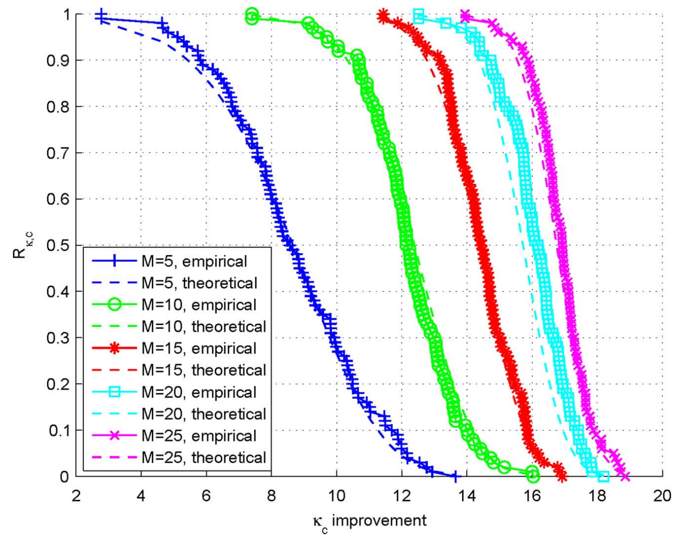


Fig. 6. Reliability function of SIR, $R_{\kappa,c}$, versus SIR improvement with various numbers of microphones.

100 microphone locations are uniformly randomized, where the number of microphones is taken from $M = 5, 10, 15, 20, 25$. As before, in each Monte-Carlo experiment and per each frequency, the SDW-MWF with $\mu = 1$ is calculated, and its SIR and white noise gain are recorded. The normalized errors of the average SIR and white noise gain, i.e., κ_c and ξ_c, are -20 dB for all tested numbers of microphones. The theoretical relation between the number of microphones, M, and the average SIR and white noise gain (50a), (50b), are verified, as the normalized errors are considerably small.

The reliability of the SIR at a frequency of 2 kHz, i.e., κ_c versus the SIR improvement for various numbers of microphones is depicted in Fig. 6. It is clear from this figure, that the derived reliability function fits the empirical data. It is interesting to note that as the number of microphones increases the reliability function converges to a step function, and hence the performance level becomes more deterministic. Similar results are obtained for other frequencies and sources constellations. As discussed earlier, the reliability measures (65a), (65b) equal the probability that the performance criteria will meet a predefined level.

B. Diffuse Sound Field

Here, we perform an experiment to verify the theoretical model of κ_{dif} and ξ_{dif} for the case of a diffuse sound field. The room dimensions, the sampling rate and the DFT size are as in the previous section, 4 m $\times$ 4 m $\times$ 3 m, 8 kHz and 8192, respectively. The number of microphones is set to $M = 16$. The reverberation time is set to one of the values 0.2 sec, 0.3 sec, $\ldots$, 0.6 sec. The received microphone signals comprise of a coherent desired source, modeled as before by an AR(6) random process, a diffuse sound field and sensors noise. As in the previous experiment, we simulate the spectra of the signals, and substitute them in the derived formulas. The average SNR of the desired source and the average INR of the diffuse sound field are set to 90 dB and 60 dB, respectively. The location of the desired source is randomly selected in 4 scenarios. For each scenario, 100 microphones positions are drawn with a 3D uniform distribution. In each Monte-Carlo experiment, and per each frequency, the SDW-MWF with $\mu = 1$ is calculated, and its SIR and white noise gain are recorded. The normalized errors of the average SIR and white noise gain, i.e., κ_{dif} and ξ_{dif}, defined as $((\langle\kappa_{\mathrm{dif}}\rangle - \bar{\kappa}_{\mathrm{dif}})^2)/(\bar{\kappa}^2_{\mathrm{dif}})$ and $((\langle\xi_{\mathrm{dif}}\rangle - \bar{\xi}_{\mathrm{dif}})^2)/(\bar{\xi}^2_{\mathrm{dif}})$, respectively, for all tested reverberation times is about -20 dB. Evidently, the formulas for the average criteria (64a), (64b) are valid.

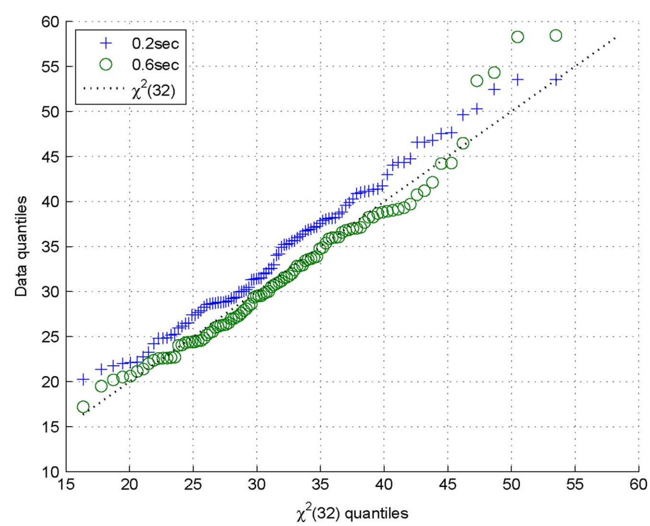


Fig. 7. Quantile-quantile probability plot of $(2\delta_{\mathrm{dif}})/(\alpha\delta_d)\kappa_{\mathrm{dif}}$ versus $\chi^2(32)$ distribution with a diffuse sound field for various reverberation times.

The following figures correspond to one of the source location scenarios at frequency 2 kHz, however, similar results are obtained at other scenarios and frequencies. In the derivation of the theoretical model, we argued that κ_{dif} is a scaled $\chi^2(2M)$ RV. The quantile-quantile probability plots of $(2\delta_{\mathrm{dif}})/(\alpha\delta_d)\kappa_{\mathrm{dif}}$ versus the $\chi^2(32)$ distribution is depicted in Fig. 7 for reverberation times of 0.2 sec, 0.6 sec. From this figure, the Chi-square distribution with $2M$ degrees of freedom of the scaled κ_{dif} can be verified. The theoretical model is verified also for other reverberation times. We also verify that $\boldsymbol{\rho}_{\mathrm{dif}}$ is an $M \times 1$ complex normal random vector. The reliability function of the SIR, i.e., $R_{\kappa,\mathrm{dif}}$, versus the SIR improvement is depicted in Fig. 8. Clearly from this figure, the reliability

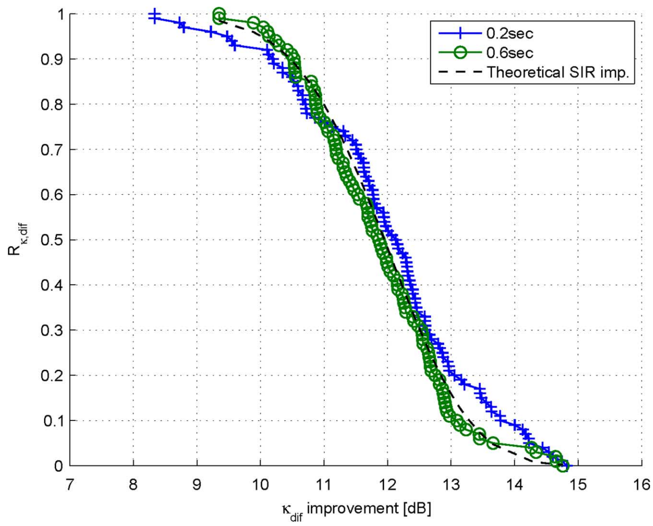


Fig. 8. Reliability function of SIR, $R_{\kappa,\mathrm{dif}}$, versus SIR improvement with a diffuse sound field for various reverberation times.

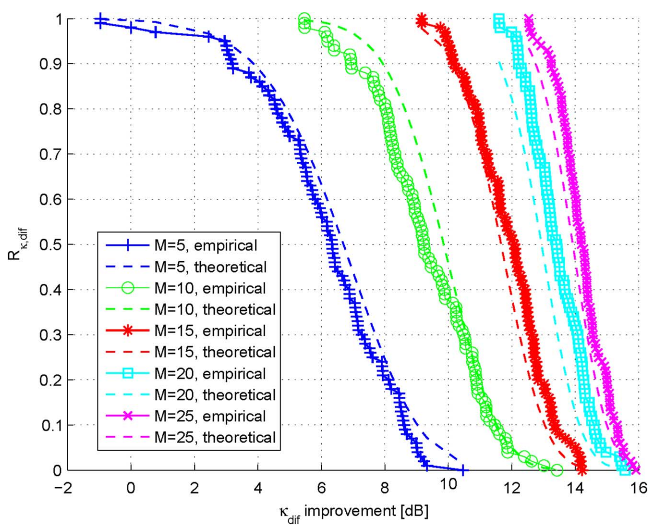


Fig. 9. Reliability function of SIR, $R_{\kappa,\mathrm{dif}}$, versus SIR improvement with various numbers of microphones.

function of the SIR is verified. Similar results were obtained in all other tested reverberation times. The theoretical model for the reliability of the white noise gain is also verified in this simulation.

The reliability functions of the SIR and white noise gain, i.e., $R_{c,\mathrm{dif}}$ and $R_{\xi,\mathrm{dif}}$, were measured with different reverberation times $T_{60} = 0.2, 0.4, 0.6, 0.8$ sec and for various values of $\mu = 1, 10, 100$. Correspondingly to derivation in (65a), (65b), the reliability criteria are independent of the parameter μ.

Now, we wish to verify the effect of the number of microphones M on the reliability measures of the SIR and white noise gain in the diffuse sound field case. We use the same room dimensions as above, and set the reverberation time to $T_{60} = 0.4$ sec. We test 4 different locations for the desired source. For each case 100 microphones locations are uniformly randomized, where the number of microphones is taken from $M = 5, 10, 15, 20, 25$. As before, in each Monte-Carlo experiment and per each frequency, the SDW-MWF with $\mu = 1$ is calculated, and its SIR and white noise gain are recorded. The normalized errors of the average SIR and white noise gain, i.e., κ_{dif} and ξ_{dif}, for all tested numbers of microphones, are about -20 dB. The theoretical relation between the number of microphones, M, and the average SIR and white noise gain (64a), (64b), are verified from these results, as the normalized errors are considerably small.

The reliability of the SIR at a frequency of 2 kHz, i.e., κ_{dif}, versus the SIR improvement for various numbers of microphones is depicted in Fig. 9. Clearly from this figure, the derived reliability function fits the empirical data. Similar results are obtained for other frequencies and source locations. As in the case of coherent noise sources, the performance tends to become deterministic as the number of microphones increases.

VII. Conclusion

We have considered the problem of signal enhancement in WASN applications where the microphone locations cannot be determined in advance. Assuming that the microphones are randomly located with a uniform distribution, and utilizing results from statistical room acoustics, we analyzed the performance of applying the SDW-MWF. Two noise fields were discussed: first, $P < M$ coherent noise sources and second, a diffuse sound field. Statistical models for two performance criteria, namely the SIR and the white noise gain, were derived for the different noise fields. Reliability functions, which give the probability of a BF criterion to exceed a predefined level, were derived for both criteria and both noise fields. The reliability functions can be used to predict the BF performance measures in a WASN, and to calculate the number of microphones needed to maintain a desired level thereof with a predefined probability. The proposed statistical models and reliability functions were verified in a comprehensive simulative study.

References

[1] D. Estrin, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2001, pp. 2033–2036.

[2] D. Culler, D. Estrin, and M. Srivastava, "Overview of sensor networks," *Computer*, vol. 37, no. 8, pp. 41–49, Aug. 2004.

[3] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks—Part I: Sequential node updating," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5277–5291, Oct. 2010.

[4] S. Doclo, M. Moonen, T. V. den Bogaert, and J. Wouters, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 1, pp. 38–51, Jan. 2009.

[5] S. Markovich-Golan, S. Gannot, and I. Cohen, "A reduced bandwidth binaural MVDR beamformer," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.

[6] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *Proc. IEEE Symp. Commun. Veh. Technol. (SCVT)*, Ghent, Belgium, Nov. 2011, pp. 1–6.

[7] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[8] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Speech distortion weighted multichannel Wiener filtering techniques for noise reduction," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. New York, NY, USA: Springer, 2005, pp. 199–228.

[9] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of multichannel Wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 4, pp. 487–503, Jul. 2005.

[10] Y. T. Lo, "A mathematical theory of antenna arrays with randomly spaced elements," *IEEE Trans. Antennas Propag.*, vol. AP-12, no. 3, pp. 257–268, May 1964.

[11] H. Ochiai, P. Mitran, H. Poor, and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4110–4124, Nov. 2005.

[12] M. F. A. Ahmed and S. A. Vorobyov, "Performance characteristics of collaborative beamforming for wireless sensor networks with Gaussian distributed sensor nodes," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2009, pp. 3249–3252.

[13] J. Huang, P. Wang, and Q. Wan, "Collaborative beamforming for wireless sensor networks with arbitrary distributed sensors," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1118–1120, Jul. 2012.

[14] K. Hardwick, D. Goeckel, D. Towsley, K. Leung, and Z. Ding, "Antenna beam pattern model for cooperative ad-hoc networks," in *Proc. ACITA*, 2008, pp. 209–216.

[15] K. C. Kerby and J. T. Bernard, "Sidelobe level and wideband behavior of arrays of random subarrays," *IEEE Trans. Antennas Propagat.*, vol. 54, no. 8, pp. 2253–2262, Aug. 2006.

[16] E. Jan and J. Flanagan, "Sound capture from spatial volumes: Matched-filter processing of microphone arrays having randomly-distributed sensors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 1996, vol. 2, pp. 917–920.

[17] I. Kodrasi, T. Rohdenburg, and S. Doclo, "Microphone position optimization for planar superdirective beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2011, pp. 109–112.

[18] S. Markovich-Golan, S. Gannot, and I. Cohen, "Performance analysis of a randomly spaced wireless microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2011, pp. 121–124.

[19] M. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *J. Audio Eng. Soc.*, vol. 35, no. 5, pp. 299–306, 1987.

[20] J. Polack, "Playing billiards in the concert hall: The mathematical foundations of geometrical room acoustics," *Appl. Acoust.*, vol. 38, no. 24, pp. 235–244, 1993.

[21] M. Schroeder, "Frequency correlation functions of frequency responses in rooms," *J. Acoust. Soc. Amer.*, vol. 34, no. 12, pp. 1819–1823, 1962.

[22] H. Kuttruff, *Room Acoustics*, 4th ed. London, U.K.: Spon, 1999.

[23] J. Jot, L. Cerveau, and O. Warusfel, "Analysis and synthesis of room reverberation based on a statistical time-frequency model," in *Proc. Audio Eng. Soc. Conv. 103*, Sep. 1997.

[24] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.

[25] J. Bitzer, K.-D. Kammeyer, and K. Simmer, "An alternative implementation of the superdirective beamformer," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust.*, New Paltz, NY, USA, Oct. 1999, pp. 7–10.

[26] F. Jacobsen and T. Roisin, "The cohrerence of reverberant sound fields," *J. Acoust. Soc. Amer.*, vol. 108, no. 1, pp. 204–210, 2000.

[27] E. Habets, "Room impulse response (RIR) generator," Sep. 2010, [Online]. Available: http://home.tiscali.nl/ehabets/rir_generator.html

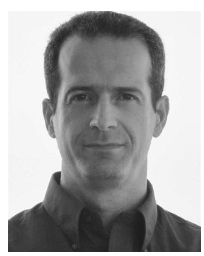
[28] N. Dal-Degan and C. Prati, "Acoustic noise analysis and speech enhancement techniques for mobile radio applications," *Signal Process.*, vol. 18, pp. 43–56, 1988.

[29] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, 2002, ch. Arrays and Spatial Filters, pp. 17–89.

Shmulik Markovich-Golan (M'12) received the B.Sc. (Cum Laude) and M.Sc. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 2002 and 2008 respectively. He is currently pursuing the Ph.D. degree at the Engineering Faculty in Bar-Ilan University. His research interests include multi-channel signal processing, distributed sensor networks, speech enhancement using microphone arrays and distributed estimation.

Sharon Gannot (S'92–M'01–SM'06) received his B.Sc. degree (summa cum laude) from the Technion Israel Institute of Technology, Haifa, Israel in 1986 and the M.Sc. (cum laude) and Ph.D. degrees from Tel-Aviv University, Israel in 1995 and 2000 respectively, all in electrical engineering. In 2001 he held a post-doctoral position at the department of Electrical Engineering (ESAT-SISTA) at K. U. Leuven, Belgium. From 2002 to 2003 he held a research and teaching position at the Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. Currently, he is an Associate Professor at the Faculty of Engineering, Bar-Ilan University, Israel, where he is heading the Speech and Signal Processing laboratory. Prof. Gannot is the recipient of Bar-Ilan University outstanding lecturer award for 2010.

Prof. Gannot is currently an Associate Editor of IEEE TRANSACTIONS ON SPEECH, AUDIO AND LANGUAGE PROCESSING. In 2003–2012 he served as an Associate Editor of the EURASIP Journal of Advances in Signal Processing, and as an Editor of two special issues on Multi-microphone Speech Processing of the same journal. He has also served as a guest editor of ELSEVIER Speech Communication journal and serves as a reviewer of many IEEE journals and conferences. Prof. Gannot is a member of the Audio and Acoustic Signal Processing (AASP) technical committee of the IEEE since Jan., 2010. He is also a member of the Technical and Steering committee of the International Workshop on Acoustic Signal Enhancement (IWAENC) since 2005 and was the general co-chair of IWAENC held at Tel-Aviv, Israel in August 2010. Prof. Gannot will serve as the general co-chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in 2013. Prof. Gannot was selected (with colleagues) to present a tutorial sessions in ICASSP 2012, EUSIPCO 2012, ICASSP 2013 and EUSIPCO 2013. Prof. Gannot research interests include multi-microphone speech processing and specifically distributed algorithms for ad hoc microphone arrays for noise reduction and speaker separation; dereverberation; single microphone speech enhancement and speaker localization and tracking.



Israel Cohen (M'01–SM'03) is a Professor of electrical engineering at the Technion—Israel Institute of Technology, Haifa, Israel. He received the B.Sc. (Summa Cum Laude), M.Sc. and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1990, 1993 and 1998, respectively.

From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT. In 2001 he joined the Electrical Engineering Department of the Technion. His research interests are statistical signal processing, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification and adaptive filtering. He is a coeditor of the Multichannel Speech Processing section of the *Springer Handbook of Speech Processing* (Springer, 2008), a coauthor of *Noise Reduction in Speech Processing* (Springer, 2009), a coeditor of *Speech Processing in Modern Communication: Challenges and Perspectives* (Springer, 2010), and a general co-chair of the 2010 International Workshop on Acoustic Echo and Noise Control (IWAENC).

Prof. Cohen is a recipient of the Alexander Goldberg Prize for Excellence in Research, and the Muriel and David Jacknow award for Excellence in Teaching. He serves as a member of the IEEE Audio and Acoustic Signal Processing Technical Committee (AASP TC) and the IEEE Speech and Language Processing Technical Committee (SLTC). He served as Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and IEEE SIGNAL PROCESSING LETTERS, and as Guest Editor of a special issue of the *EURASIP Journal on Advances in Signal Processing* on Advances in Multi-microphone Speech Processing and a special issue of the *Elsevier Speech Communication Journal* on Speech Enhancement.