PERFORMANCE ANALYSIS OF A RANDOMLY SPACED WIRELESS MICROPHONE ARRAY

Shmulik Markovich Golan[1], *Sharon Gannot*[1] *and Israel Cohen*[2]

[1] School of Engineering
Bar-Ilan University
Ramat-Gan, 52900, Israel
shmulik.markovich@gmail.com; gannot@eng.biu.ac.il

[2] Department of Electrical Engineering
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel
icohen@ee.technion.ac.il

ABSTRACT

A randomly distributed microphone array is considered in this work. In many applications exact design of the array is impractical. The performance of these arrays, characterized by a large number of microphones deployed in vast areas, cannot be analyzed by traditional deterministic methods. We therefore derive a novel statistical model for performance analysis of the MWF beamformer. We consider the scenario of one desired source and one interfering source arriving from the far-field and impinging on a uniformly distributed linear array. A theoretical model for the MMSE is developed and verified by simulations. The applicability of the proposed statistical model for speech signals is discussed.

***Index Terms*—** Random arrays, Beamforming

1. INTRODUCTION

The concept of distributed sensor networks is becoming more realistic with technology advances in the fields of nano-technology, micro electro-mechanic systems (MEMS) and communication. A distributed sensor network comprises scattered nodes which are autonomous, self-powered modules consisting of sensors, actuators and a transceiver. Their layout and connectivity graph are usually random and dynamic. Distributed sensor networks have a broad range of applications which can be categorized in ecology, environment monitoring, medical, security and surveillance.

Consider the border security application as a test case. In this application, microphones are deployed along the border and are used for detecting intruders, and for eavesdropping. The desired signal is usually contaminated by environmental noises necessitating the use of noise reduction techniques. The limited performance of single channel noise reduction algorithms, and the inherent multiple nodes structure call upon incorporating array processing algorithms.

Van Veen and Buckley [1] analyzed the performance of various beamformers. Specifically, they investigated the influence of physical properties such as the array aperture and the sensors spacing on the beam-pattern. They showed that data dependent beamformers such as the multi-channel Wiener filter (MWF) and the minimum variance distortionless response (MVDR) beamformer outperform data independent beamformers as the delay and sum (DS) beamformer. Doclo and Moonen [2] applied the MWF beamformer for speech processing, and proposed practical methods for estimating the required statistics. Both of these contributions analyze properties of the beamformers by utilizing the array layout. As we consider random arrays in the current work, such an analysis is inadequate.

Incorporating statistical models into the sensors' spatial distribution was proposed in the past by Lo [3], yielding better understanding of the beampattern properties (e.g. directivity, beam-width and sidelobe level) of a simple DS beamformer applied in a linear array. Lo treated the array properties as random variables (RVs), which are functions of the random array constellation and of the sources configuration. Ochiai et al. [4], and Ahmed and Vorobyov [5] extended the discussion, and examined the simple DS beamformer with planar random arrays in a three dimensional space. Jan and Flanagan [6] proposed to use a data dependent matched filter beamformer in a distributed microphone array, and presented its performance in a reverberant environment. Yet, no theoretical analysis on the performance of data dependent beamformers for random layouts has been made thus far.

In the current contribution a statistical model for the performance of the MWF beamformer is derived for a randomly located linear microphone array in a typical speech enhancement scenario. As stated earlier, the MWF is considered here since data-dependent beamformers are more suitable to speech processing than their data-independent counterparts [2]. We treat the scenario of a coherent wide-band desired source and a coherent wide-band interfering source arriving from the far-field and impinging on the microphone array, in a non-reverberant environment. Other scenarios can be treated in a similar fashion.

In Sec. 2, the problem is formulated. In Sec. 3, a formula for the minimum mean squared error (MMSE) of the MWF, given the microphones locations, is derived. Then, in Secs. 4 and 5 the statistics of the MMSE cost function is analyzed. The derived theoretical models are verified in Sec. 6. Aspects of applying the MWF in randomly distributed microphone arrays to speech processing are discussed in Sec. 7.

2. PROBLEM FORMULATION

Consider a coherent wideband desired source and a coherent wideband interfering source impinging on a linear array of randomly spaced microphones from the far-field in a reverberant-free environment. The array is assumed to comprise M uniformly distributed microphones in the range $\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$, where Δ is the array aperture. The microphones locations are denoted by $x_1, \ldots, x_M$. In the short time Fourier transform (STFT) domain, the desired source is denoted by $s_d(\ell, k)$, and the interfering source is denoted by $s_i(\ell, k)$, where ℓ is the frame index, and k is the frequency index. The analysis window length is denoted by N_{DFT}. For simplicity, the desired source is assumed to arrive from the broadside. The angle between the interfering and the desired sources is denoted by θ_i. The received signals are denoted in vector notation by:

$$\mathbf{z}(\ell,k)=\sqrt{M}\mathbf{h}_d(\ell,k)s_d(\ell,k)+\sqrt{M}\mathbf{h}_i(\ell,k)s_i(\ell,k)+\mathbf{u}(\ell,k) \tag{1}$$

where $\mathbf{h}_d(\ell,k)$ and $\mathbf{h}_i(\ell,k)$ are the normalized acoustic transfer functions (ATFs) relating the desired and interfering sources and the microphones, respectively, and $\mathbf{u}(\ell,k)$ is a spatially-white sensors noise. The kth wavelength corresponding to the kth frequency index is $\lambda_k=\frac{2\pi c}{f_s}\frac{N_{\text{DFT}}}{k}$ where f_s is the sampling frequency and c is the sound velocity.

The desired and interfering ATFs are assumed time-invariant and are given by:

$$\mathbf{h}_d(\ell,k)=\frac{1}{\sqrt{M}}\left[\underbrace{\begin{matrix}1 & \cdots & 1\end{matrix}}_{M}\right]^T \tag{2a}$$

$$\mathbf{h}_i(\ell,k)=\frac{1}{\sqrt{M}}\left[\begin{matrix}e^{-j\xi_i\frac{x_1}{\lambda_k}} & \cdots & e^{-j\xi_i\frac{x_M}{\lambda_k}}\end{matrix}\right]^T \tag{2b}$$

where $\xi_i=2\pi\sin(\theta_i)$. Note that by this notation the phase of both the desired source and of the interfering source is assumed to be 0 at the origin ($x=0$). Note also that unlike the common notation, the ATFs in our work are normalized. For brevity, the frequency index is omitted and only the kth frequency index is considered. The same analysis is applicable to each frequency bin. Define the total interference by $\mathbf{v}(\ell)=\sqrt{M}\mathbf{h}_i s_i(\ell)+\mathbf{u}(\ell)$. The second moments of the received signals are denoted by:

$$\mathbf{\Phi}_{zz}=M\sigma_d^2\mathbf{h}_d{\mathbf{h}_d}^\dagger+\mathbf{\Phi}_{vv} \tag{3}$$

where

$$\mathbf{\Phi}_{vv}=M\sigma_i^2\mathbf{h}_i{\mathbf{h}_i}^\dagger+\sigma_u^2\mathbf{I} \tag{4}$$

is the covariance matrix of the total interference, $\sigma_d^2,\sigma_i^2,\sigma_u^2$ are the spectra in the kth frequency bin of the desired source, the interfering source and the microphone noise, respectively, and $\mathbf{I}$ is the $M\times M$ identity matrix. The goal of the MWF is to estimate a delayed version of the desired signal in the MMSE sense. Here, the desired signal is defined by the desired source component at the first microphone:

$$d(\ell)=\sqrt{M}h_{d,1}s_d(\ell). \tag{5}$$

The output of the beamformer $\mathbf{w}$ is denoted by $y_o(\ell)=\mathbf{w}^\dagger\mathbf{z}(\ell)$.

3. MSE ANALYSIS GIVEN THE MICROPHONE LOCATIONS

As the microphone locations are random, the corresponding MWF and its corresponding MMSE are also RVs. Their statistics is analyzed in the following sections. In this section the mean squared error (MSE) of the MWF is analyzed for a given set of microphone locations. The MWF is given by:

$$\mathbf{w}_{\text{MWF}}=\mathbf{\Phi}_{zz}^{-1}\boldsymbol{\phi}_{zd} \tag{6}$$

where

$$\boldsymbol{\phi}_{zd}=\mathrm{E}\left[\mathbf{z}(\ell)d(\ell)^*\right] \tag{7}$$

is the cross-correlation vector between the received signals and the desired signal. Substituting (1) in (7) gives

$$\boldsymbol{\phi}_{zd}=M\sigma_d^2h_{d,1}\mathbf{h}_d. \tag{8}$$

For further simplification of (6), the Woodbury identity is applied to $\mathbf{\Phi}_{zz}$ in (3):

$$\begin{aligned}\mathbf{\Phi}_{zz}^{-1}&=\mathbf{\Phi}_{vv}^{-1}-\mathbf{\Phi}_{vv}^{-1}\mathbf{h}_d\left(\left(M\sigma_d^2\right)^{-1}+{\mathbf{h}_d}^\dagger\mathbf{\Phi}_{vv}^{-1}\mathbf{h}_d\right)^{-1}{\mathbf{h}_d}^\dagger\mathbf{\Phi}_{vv}^{-1}\\&=\left(\mathbf{I}-\left(\left(M\sigma_d^2\right)^{-1}+\alpha\right)^{-1}\mathbf{\Phi}_{vv}^{-1}\mathbf{h}_d{\mathbf{h}_d}^\dagger\right)\mathbf{\Phi}_{vv}^{-1}\end{aligned} \tag{9}$$

where

$$\alpha={\mathbf{h}_d}^\dagger\mathbf{\Phi}_{vv}^{-1}\mathbf{h}_d. \tag{10}$$

Substituting (9) and (8) in (6) yields the expression

$$\mathbf{w}_{\text{MWF}}=\frac{h_{d,1}}{\left(M\sigma_d^2\right)^{-1}+\alpha}\mathbf{\Phi}_{vv}^{-1}\mathbf{h}_d. \tag{11}$$

The corresponding MMSE is given by:

$$J_{\text{MWF}}=\mathrm{E}\left[\left|d(\ell)-{\mathbf{w}_{\text{MWF}}}^\dagger\mathbf{z}(\ell)\right|^2\right]=\sigma_d^2-{\mathbf{w}_{\text{MWF}}}^\dagger\mathbf{\Phi}_{zz}\mathbf{w}_{\text{MWF}}. \tag{12}$$

Substituting (11) in (12) gives

$$J_{\text{MWF}}=\sigma_d^2\frac{\left(M\sigma_d^2\right)^{-1}}{\left(M\sigma_d^2\right)^{-1}+\alpha}. \tag{13}$$

Denote the inner product of the desired and interfering ATFs by:

$$\rho={\mathbf{h}_d}^\dagger\mathbf{h}_i. \tag{14}$$

Applying the Woodbury identity to $\mathbf{\Phi}_{vv}$ in (4) and substituting the result in (10) yields

$$\alpha=\frac{1}{\sigma_u^2}\left(1-\frac{\sigma_i^2}{\sigma_i^2+\sigma_u^2/M}|\rho|^2\right). \tag{15}$$

Finally, denoting the following spectra ratios

$$\gamma_d=\frac{\sigma_d^2}{\sigma_d^2+\sigma_u^2/M} \tag{16a}$$

$$\gamma_i=\frac{\sigma_i^2}{\sigma_i^2+\sigma_u^2/M} \tag{16b}$$

and substituting (15) in (13) gives

$$J_{\text{MWF}}\left(|\rho|^2\right)=\frac{\sigma_u^2}{M}\gamma_d\left(1-\gamma_d\gamma_i|\rho|^2\right)^{-1}. \tag{17}$$

The last expression for the MMSE depends only on the sources spectra, the number of microphones, and ρ. The residual error power, J_{MWF}, is a monotonically increasing function of $|\rho|^2$. Considering that $0\le|\rho|^2\le 1$ and the monotonic behavior of J_{MWF}, the range of J_{MWF} is bounded by $\sigma_d^2\frac{\sigma_u^2/M}{\sigma_d^2+\sigma_u^2/M}\le J_{\text{MWF}}\le\sigma_d^2\frac{\sigma_i^2+\sigma_u^2/M}{\sigma_d^2+\sigma_i^2+\sigma_u^2/M}$. The lower bound corresponds to the case where the sources' ATFs are orthogonal. Its corresponding MWF is the average of M identical single channel Wiener filters calculated at the absence of the interference. The upper bound corresponds to the case where the sources' ATFs coincide. In this case the residual error power is maximized since it is impossible to spatially separate the desired and interfering sources. Its corresponding MWF is the average of M identical single channel Wiener filters assuming that the interference is present.

In Sec. 4 the microphone locations are assumed random and the statistics of ρ is analyzed. The statistics of ρ will then be used for analyzing the statistics of J_{MWF} based on (17) .

4. THE STATISTICS OF ρ

A summary of the statistics of ρ which was derived by Lo in [3] follows. ρ is a complex RV, $\rho = \rho_r + \mathrm{i}\rho_i$, with real and imaginary components denoted by ρ_r and ρ_i. Using (14) and (2a,2b) it can be verified that

$$\rho_r = \frac{1}{M}\sum_{m=1}^{M}\cos\left(\xi_i\frac{x_m}{\lambda}\right) \quad (18a)$$

$$\rho_i = \frac{1}{M}\sum_{m=1}^{M}\sin\left(\xi_i\frac{x_m}{\lambda}\right) \quad (18b)$$

where $\mathrm{i} = \sqrt{-1}$. Now since $\{x_m\}_{m=1}^{M}$ are independent identically distributed (i.i.d.) RVs, it can be shown that the first-order and second-order moments of ρ_r and ρ_i are given by:

$$\mu_{\rho,r} = \mathrm{E}\left[\rho_r\right] = \phi_x(\frac{\xi_i}{\lambda}) \quad (19a)$$

$$\sigma^2_{\rho,r} = \mathrm{E}\left[(\rho_r - \mu_{\rho,r})^2\right] = \frac{1}{2M}\left(1 + \phi_x(2\frac{\xi_i}{\lambda})\right) - {\phi_x}^2(\frac{\xi_i}{\lambda}) \quad (19b)$$

$$\mu_{\rho,i} = \mathrm{E}\left[\rho_i\right] = 0 \quad (19c)$$

$$\sigma^2_{\rho,i} = \mathrm{E}\left[(\rho_i - \mu_{\rho,i})^2\right] = \frac{1}{2M}\left(1 - \phi_x(2\frac{\xi_i}{\lambda})\right) - {\phi_x}^2(\frac{\xi_i}{\lambda}) \quad (19d)$$

where $\phi_x(t) = \mathrm{sinc}(t\Delta)$ denotes the characteristic function of the RV x, in the Uniform distribution case. The summands of the summation in ρ_r, ρ_i are i.i.d. RVs. Therefore, according to the central limit theorem (CLT) they converge to a Gaussian RV for $M \gg 1$. Assuming that

$$\frac{\xi_i\Delta}{\lambda} \gg 1, \quad (20)$$

the following approximation holds:

$$\rho_r \sim \mathcal{N}(0, \frac{1}{2M}) \quad (21a)$$

$$\rho_i \sim \mathcal{N}(0, \frac{1}{2M}). \quad (21b)$$

Note that this approximation is not valid for $\theta_i \to 0$. Since ρ_r and ρ_i are uncorrelated, i.e. $\mathrm{E}\left[\rho_r\rho_i\right] = 0$, $2M|\rho|^2 = 2M{\rho_r}^2 + 2M{\rho_i}^2$ is approximated by a χ^2 RV with 2 degrees of freedom, which is an Exponential RV with a parameter $1/2$, i.e.

$$2M|\rho|^2 \sim \exp(1/2). \quad (22)$$

Note that when assumption (20) holds, ρ tends to its lower bound 0 as the number of microphones M increases, and according to (17) the corresponding J_{MWF} tends to its lower bound as well.

5. STATISTICS OF THE MSE

Using the Exponential distribution of $|\rho|^2$, the expression for J_{MWF} in (17) and its monotonic behaviour, the cumulative distribution function (CDF) of J_{MWF} is given by:

$$\begin{aligned}\Pr\left(J_{\mathrm{MWF}}(|\rho|^2) \le J_0\right) = \Pr\left(|\rho|^2 \le {J_{\mathrm{MWF}}}^{-1}(J_0)\right) = \\ \Pr\left(|\rho|^2 \le {\gamma_d}^{-1}{\gamma_i}^{-1}\left(1 - \frac{\sigma_u^2/M}{J_0}\gamma_d\right)\right) = \\ 1 - \exp\left(-M{\gamma_d}^{-1}{\gamma_i}^{-1}\left(1 - \frac{\sigma_u^2/M}{J_0}\gamma_d\right)\right).\end{aligned} \quad (23)$$

Equation (23) denotes a reliability measure of J_{MWF}. It equals the probability that the MMSE will not exceed a desired level, J_0.

6. MODEL VERIFICATION

We turn now to the verification of the derived models. We verify the Normal model of ρ in (21a,21b), and the Exponential models of $|\rho|^2$ and the reliability function in (22) and (23), respectively. For each scenario a Monte-Carlo simulation consisting of 1000 arrays of $M = 21$ microphones (unless stated otherwise) were randomized with a Uniform distribution on a linear aperture of $\Delta = 10$ length units. A desired source and an interfering source arriving from the far-field were simulated. The angle of arrival (AOA) of the desired source was set to 0^o, and the AOA of the interference was set to $\theta_i = 5.5^o$ (unless stated otherwise). A low level sensors noise was added to the received signals. The signal to interference ratio (SIR) was set to 0dB and the signal to noise ratio (SNR) was set to 30dB. The results were obtained for wavelengths in the range of $[0.1, 2]$ length units, and are shown for a a specific wavelength of $\lambda = 0.91$ length units (unless stated otherwise).

6.1. The Normal model of the components of ρ and the Exponential model of $|\rho|^2$

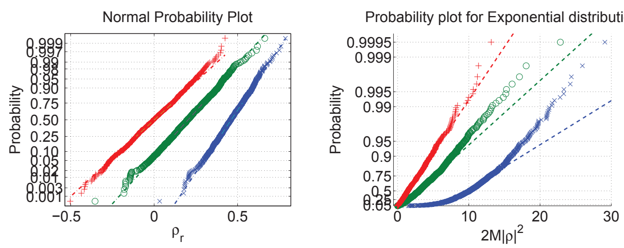
The Normal probability plots of ρ_r and the Exponential probability plots of $2M|\rho|^2$ were simulated for various combinations of λ and θ_i and the results are shown in Fig. 1(a) and Fig. 1(b), respectively. The blue color corresponds to $\theta_i = 1.5^o, \lambda = 0.41 \Rightarrow \xi_i\frac{\Delta}{\lambda} = 0.64$, the green color corresponds to $\theta_i = 9.5^o, \lambda = 2 \Rightarrow \xi_i\frac{\Delta}{\lambda} = 0.82$, and the red color corresponds to $\theta_i = 9.5^o, \lambda = 0.41 \Rightarrow \xi_i\frac{\Delta}{\lambda} = 4.02$. The various markers (cross, circle and plus) correspond to the measurements data and the dashed lines correspond to the best distribution fit. Markers that coincide with the dashed lines correspond to a good fit between the distribution model and the data.

Since assumption (20) holds only for the parameters of the red curve, we expect it to coincide with the theoretical distribution models in (21a,21b) and (22). It is clear from Fig. 1(a) that a Normal distribution model fits the blue, green and red curves. However, only the red curve matches the zero mean model of ρ in (21a,21b). From Fig. 1(b) it is obvious that only the red curve matches the Exponential distribution. The estimated mean and variance of ρ_r are -3.1×10^{-3} and 2.34×10^{-2}. These values match the theoretical values of 0 and $2.38 \times 10^{-2} = \frac{1}{2\times 21}$ in (21a,21b). The estimated parameter of the Exponential distribution of $2M|\rho|^2$ is 0.49. It matches its theoretical value of 0.5 in (22). The blue and the green curves are examples for cases where assumption (20) is invalid, rendering the Exponential distribution inappropriate for representing the data points.

6.2. The reliability of J_{MWF}

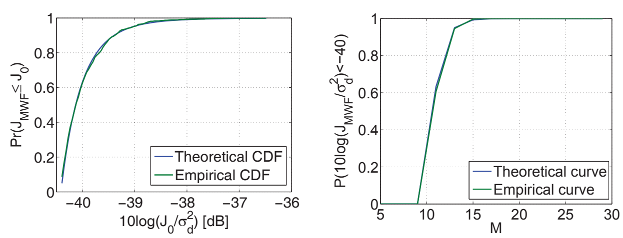
In order to verify the reliability of J_{MWF} the number of microphones was set in the range $M = 5, 6, \ldots, 30$. The analytical and empirical reliability functions for $M = 11$ microphones are depicted in Fig. 2(a). It is clear from this figure that the theoretical model in (23) matches the empirical data. The MMSE normalized by the signal power in this case varies in the range of -41dB to -36dB.

The theoretical value of the CDF of the MMSE at point $\Pr\left(10\log\left(\frac{J_{\mathrm{MWF}}}{\sigma_d^2}\right) \le -40\right)$, as well as its empirical value are depicted in Fig. 2(b). It is clear from this figure that the theoretical model matches the empirical results. It is evident that the CDF has an approximate step function characteristics. Below a certain threshold (number of microphones) the CDF tends to 0 and above the threshold it tends to 1 with an abrupt transition between the two values. The threshold in Fig. 2(b) is approximately at $M = 11$ microphones. Using more than 11 microphones, it is almost guaranteed that the MWF error will be lower than -40dB.



(a) Normal probability plots of ρ_r (b) Exponential probability plots of $2M|\rho|^2$

Fig. 1. Normal and Exponential probability plots of ρ_r and $|\rho|^2$, respectively, for various values of $\xi_i \overset{\Delta}{=} \frac{}{\lambda}$ (0.64 in blue, 0.82 in green and 4.02 in red).



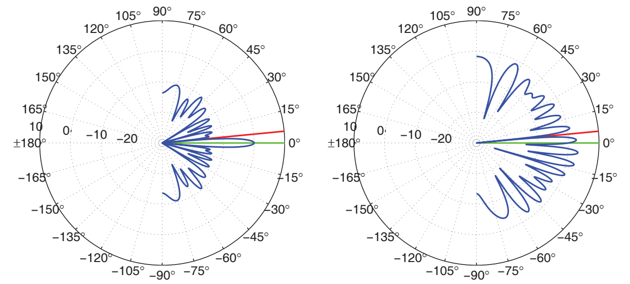
(a) Theoretical (blue) and empirical (green) CDF curves of J_{MWF} with an $M = 11$ microphones array (b) Theoretical (blue) and empirical (green) values of $\Pr\left(10\log\left(\frac{J_{\mathrm{MWF}}}{\sigma_d^2}\right) \leq -40\right)$ vs. M, the number of microphones

Fig. 2. The CDF of J_{MWF} as a function of J_0 and M, the number of microphones

Examples for a beampattern corresponding to an array realization with $M = 9$ microphones (below the threshold), and for a beampattern corresponding to an array realization with $M = 15$ microphones (above the threshold) are depicted in Fig. 3(a) and Fig. 3(b), respectively. The beampattern in Fig. 3(a) has low sidelobes while the beampattern in Fig. 3(b) seems to suffer from spatial aliasing and as a results, exhibits high sidelobes.

7. CONCLUSIONS

In many applications, e.g. border security, exact design of the array is impractical. The performance of these arrays, characterized by a large number of microphones deployed in vast areas, cannot be analyzed by traditional deterministic methods. In the current contribution we have presented a statistical model for the performance of a randomly located linear microphone array. Specifically, we analyzed



(a) An example beampattern with $M = 15$ microphones. (b) An example beampattern with $M = 9$ microphones

Fig. 3. Examples for beampatterns for two array realizations. The AOAs of the desired and interfering sources are depicted in green and red, respectively.

the MMSE measure of the MWF, commonly used in speech applications. The case of one desired source and one interfering source arriving from the far-field was treated. The theoretical models that have been developed were verified by simulations. The proposed model can be used for determining the number of microphones required for obtaining a predefined residual error level.

Special considerations need to be made when processing speech signals. Throughout this work we assumed that the second-order statistics of the various sources is available. This is hardly ever the case in actual applications. A practical design of the MWF for speech processing is given by Doclo and Moonen [2]. The derivation of the reliability measure assumes that the sources are stationary with known spectra. The reliability measure can be extended to the case of speech signals in several ways. The stationary spectra can be replaced by instantaneous spectra estimates, or time averaged spectra. Alternatively, the required number of microphones can be determined by worst-case considerations.

8. REFERENCES

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

[2] S. Doclo and M. Moonen, "GSVD-Based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Sig. Proc.*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.

[3] Y. T. Lo, "A mathematical theory of antenna arrays with randomly spaced elements," *IEEE Trans. Ant. and Prop.*, vol. 12, pp. 257–268, May 1964.

[4] H. Ochiai, P. Mitran, H. V. Poor, and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks," *IEEE Trans. Sig. Proc.*, vol. 53, pp. 4110–4124, Nov. 2005.

[5] M. F. A. Ahmed and S. A. Vorobyov, "Performance characteristics of collaborative beamforming for wireless sensor networks with gaussian distributed sensor nodes," *Proc. ICASSP*, pp. 3249–3252, 2009.

[6] E. Jan and J. Flanagan, "Sound capture from spatial volumes: matched-filter processing of microphone arrays having randomly-distributed sensors," *Proc. ICASSP*, vol. 2, pp. 917–920, 1996.