https://doi.org/10.1016/j.engappai.2015.04.015

Beamformer configuration design in reverberant environments

Zhibao Li and Ka-Fai Cedric Yiu

Abstract

Many speech-related products rely on the deployment of microphone arrays and standard regular configurations are often used. In enhancing speech quality, the placement of microphones is indeed an important factor. Moreover, for indoor applications, the room acoustics further increases the difficulty. In this paper, these problems are addressed. First, we define the LCMV beamformer design problem using estimated room impulse responses with reverberation. Then, we study the performance limit on the filter length and formulate the configuration design problem. Finally, we employ a hybrid descent method with the genetic algorithm for solving the design problem. Numerical examples demonstrate the effectiveness of the proposed method.

Keywords: Microphone array configuration, reverberation, LCMV beamforming, hybrid descent method.

I. Introduction

Beamforming techniques have been employed extensively in speech communication systems, teleconferencing, speech recognition, and hearing aids [3], [14]. Beamformers act as spatial filters to extract a target from a mixture of signals captured by a set of microphones. Many beamforming techniques were developed under the assumption that channels are modelled by delays and attenuations. For instance, the minimum variance distortionless response (MVDR) beamformer originally proposed in [5] consists of minimizing the overall interference-plus-noise power subject to a gain constraint in the speaker direction. The linearly constrained minimum variance (LCMV) beamformer [9], [13] generalizes the idea with multiple constraints imposed. When the performances of these beamforming techniques are investigated, it is well-known that the filter length plays an important role; however the performance limit begins to plateau out for long enough filters. As proven in [12], unless

Zhibao Li is with the Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, PR China, e-mail: zbli0307@163.com

Ka-Fai Cedric Yiu is the corresponding author, with the Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, PR China, e-mail: macyiu@polyu.edu.hk

the number of microphones grows significantly, the target response might not be achieved satisfactorily. On the other hand, for a fixed size array, the performance can still be improved drastically if the configuration is carefully chosen [11]. Furthermore, for indoor applications, it is also necessary to match the microphone array configuration with the specific auditory scene. Indeed, in designing the products embedded with microphones, it is essential to find suitable beamforming filters to enhance the desire signal. In addition, it is a common practice to employ regular configurations in the products without considering different microphone arrangements. For example, there are products on acoustic measurements which deploy regular shapes such as planar wheel arrays, spherical arrays and rectangular arrays. They can be applied for environmental noise measurements or for indoor applications for sound reception. However, optimization of these products have not been considered in the literature for different environments.

If the microphone locations are restricted to vary within certain dimensions and areas, several optimization problems have been formulated. For example, if microphones are displaced linearly in a one-dimensional manner, it essentially reduces to the array thinning technique [24], [25], [31], [32]; different algorithms have been developed, including evolutionary programming [7], [17], genetic algorithm [6], [16], simulated annealing algorithm [8], [34], [35] and pattern search algorithm [29]. For applications inside a vehicle, microphones are restricted to be in several dedicated areas and an evolutionary algorithm has been proposed in [2]. In formulating the general multi-dimensional design problem, a nonlinear optimization problem using the L_2-norm was proposed in [11], which allows microphones to move around in a multi-dimensional solution space in search of better configurations. However, the objective function is highly nonlinear and is essentially nonconvex with respect to the location variables; the problem is further complicated by the influence of the filter length. By considering the performance limit for sufficiently long filters, a reduced optimization problem with microphone locations being the only set of decision variables was proposed [11]. Also, a hybrid descent method using the genetic algorithm was developed in [21] to provide a more general solution technique to the problem.

For indoor applications in a reverberant environment, reverberation makes the transfer function estimation for sound wave propagation more complicated and costly. Room acoustics can be estimated by geometric based models, such as the ray tracing method and the image-source method (ISM) [1], [4], [19], [33], and have been applied for indoor beamformer design [20], [22]. During the configuration design process, the room acoustics must be recalculated whenever the configuration changes. A fast implementation of the ISM proposed in [18] provides the required tool for synthesizing the room impulse responses (RIRs) in an efficient manner, requiring only a fraction of the computation time comparing with the

original approach. In this paper, the fast-ISM technique is adopted and the RIRs information is embedded into the LCMV beamformer design problem. By optimizing on the frequency response function directly to obtain the performance limit on the filter length, a pure location optimization problem is formulated. A suitable hybrid descent method with the genetic algorithm is developed to tackle the design problem.

The rest of the paper is organized as follows. In Section II, we define the configuration design problem by using the estimated RIRs based on the fast image-source method. In Section III, we study the performance limit on the filter length of the LCMV beamformer by solving the optimal frequency response function directly. In Section IV, after eliminating the filter length effect, we develop a hybrid descent method using the genetic algorithm to tackle the configuration design problem. In Section V, several numerical examples are given to evaluate the effectiveness of the proposed method. Conclusion is given in Section VI.

II. Microphone array configuration design

Assume the indoor environment is a rectangular room with N sources settled at $\boldsymbol{r}_n$, $n = 0, \ldots, N-1$. Sound wave propagation in the room is governed by the room impulse responses (RIRs) (Fig. 1) from all source points to the microphone array, which can be estimated by the image-source method.

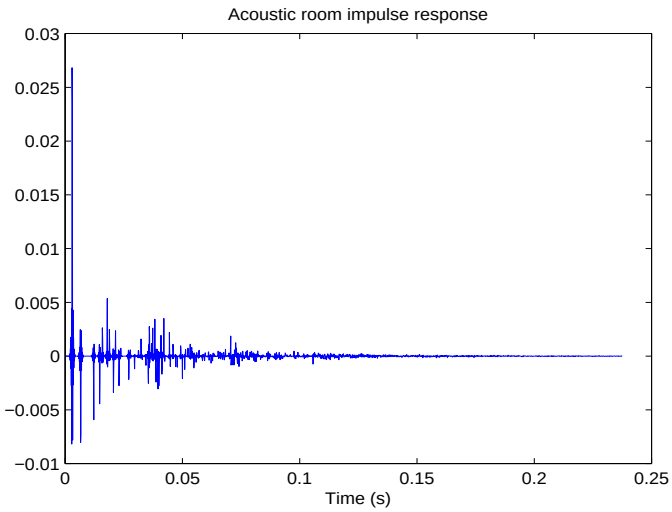


Fig. 1: A typical example of the RIR.

Without loss of generality, let $\boldsymbol{r}_0$ be the source point of the signal of interest (SOI) while the others are the source points of interferences (INT). For a given M-elements microphone array located at $\boldsymbol{\lambda} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_M) \in \boldsymbol{\Lambda} \subset \mathbb{R}^{3\times M}$, suppose the frequency domain transfer functions describing the acoustics are denoted by $H_{n,m}(\boldsymbol{\lambda}, \omega)$, signals captured by

the microphone array are represented by

$$X_m(\omega) = \sum_{n=0}^{N-1} H_{n,m}(\omega) S_n(\omega) + V_m(\omega), \quad m = 1, \ldots, M. \tag{1}$$

Assume there is an $L-$tap FIR filter behind each microphone with coefficients $w_m = [w_m(0),$ $w_m(1),$ $\ldots,$ $w_m(L-1)]^T$, $m = 1, \ldots, M$, if signals received by this microphone array are sampled synchronously at the rate of f_s per second, then the frequency responses for the frequency component ω are

$$W_m(\omega) = w_m^T d_0(\omega), \quad m = 1, \ldots, M, \tag{2}$$

where $d_0(\omega)$ is defined as

$$d_0(\omega) = [e^{\frac{-j\omega}{f_s}(-\tau_L)}, e^{\frac{-j\omega}{f_s}(1-\tau_L)}, ..., e^{\frac{-j\omega}{f_s}(L-1-\tau_L)}]^T, \tag{3}$$

and $0 \leq \tau_L \leq L-1$ is the group delay. Therefore, the beamformer output for each frequency under the array placement $\boldsymbol{\lambda}$ is given by

$$Y(\omega) = \sum_{m=1}^{M} W_m(\omega) X_m(\omega) = \boldsymbol{W}^H(\omega) \boldsymbol{X}(\omega) \tag{4}$$

where $\boldsymbol{W}(\omega) = [W_1(\omega) \ \ldots \ W_M(\omega)]^T$ denotes the beamformer response vector and $\boldsymbol{X}(\omega) = [X_1(\omega) \ \ldots \ X_M(\omega)]^T$ is the received signal.

Notice that the beamformer output $Y(\omega)$ is a function of the microphone array placement $\boldsymbol{\lambda}$ and filter coefficients $\boldsymbol{w} = [w_1, w_2, \ \ldots, \ w_M]^T$. To measure the error between the output and the desired SOI signal $S_d(\omega)$, we define a merit function as

$$\boldsymbol{F}(\boldsymbol{\lambda}, \boldsymbol{w}) = \frac{1}{||\Omega||_2} \int_{\Omega} ||\boldsymbol{W}^H(\omega) \boldsymbol{X}(\omega) - S_d(\omega)||_2^2 d\omega, \tag{5}$$

where Ω is the interesting frequency region. Hence, we can propose the microphone array configuration design problem as follows:

$$\begin{aligned} \min_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}, \boldsymbol{w} \in \mathbb{R}^{M \times L}} \quad & \boldsymbol{F}(\boldsymbol{\lambda}, \boldsymbol{w}) \\ s.t. \quad & ||\boldsymbol{r}_i - \boldsymbol{r}_j||^2 \geq \varepsilon_0, \ i, j = 1, 2, \ldots, M, \ i \neq j, \end{aligned} \tag{6}$$

where $\boldsymbol{F}(\boldsymbol{\lambda}, \boldsymbol{w})$ is defined in (5), $\varepsilon_0 > 0$ is a constant to separate the set of microphone elements for proper functioning. This design problem (6) can be considered as a bi-level optimization problem with placement variables $\boldsymbol{\lambda}$ and filter coefficients $\boldsymbol{w}$. Clearly, the optimal filter coefficients $\boldsymbol{w}$ will change whenever perturbing the locations $\boldsymbol{\lambda}$. Moreover, the performance of the beamformer output $Y(\omega)$ is affected by the filter length L. In the following, we introduce the performance limit on the filter length to estimate the optimal beamformer output $Y(\omega)$ in order to avoid the effect of filter length.

III. Performance limit of LCMV beamformer

For a given $\boldsymbol{\lambda}$, the optimal beamformer output $Y(\omega)$ will be achieved when the filter length $L \to +\infty$. While the beamformer design with very long filters are very computationally costly, there is an equivalent relationship between infinite length filters and frequency response functions according to (2). The exact frequency responses $W_{m,opt}(\omega)$, $m = 1, \ldots, M$ are associated with the infinite length filters [10] in the following lemma.

Lemma 1. *Given a space spanned by the set of functions defined from* $d_0(\omega)$ (3) *as*

$$\mathcal{B} = \{e^{\frac{-j\omega}{f_s}(-\tau_L)}, e^{\frac{-j\omega}{f_s}(1-\tau_L)}, ..., e^{\frac{-j\omega}{f_s}(L-1-\tau_L)}\}. \tag{7}$$

For any complex valued function $u(\omega) + jv(\omega) \in \mathcal{C}$*, where* $u(\omega), v(\omega) \in \mathcal{R}$*, if* $u(\omega)$ *and* $v(\omega)$ *are continuous, absolute integrable and differentiable, then this complex valued function* $u(\omega) + jv(\omega)$ *can be linear represented by the base* $\mathcal{B}$ *at* $L \to +\infty$*, that is there exists a real sequence* $\{c_l,\ l = 0, 1,\ \ldots,\ +\infty\}$ *such that*

$$u(\omega) + jv(\omega) = \lim_{L \to +\infty} \sum_{l=0}^{L-1} c_l e^{\frac{-j\omega}{f_s}(l-1).} \tag{8}$$

Proof. See [10] for the details. □

From Lemma 1, we define the frequency response $W_m(\omega) = u_m(\omega) + jv_m(\omega)$ as the variables directly to estimate the optimal beamformer output, which represents the performance limit on the filter length. By substituting the received signals $X_m(\omega)$ to different channels (1), we can rearrange the beamformer output as

$$Y(\omega) = \sum_{m=1}^{M} W_m(\omega) H_{0,m}(\omega) S_0(\omega) + \sum_{n=1}^{N-1} \sum_{m=1}^{M} W_m(\omega) H_{n,m}(\omega) S_n(\omega) + \sum_{m=1}^{M} W_m(\omega) V_m(\omega). \tag{9}$$

Therefore, the subproblem for optimal beamforming output is converted to finding the optimal frequency responses $W_{m,opt}(\omega)$, such that the first part of the above sum reconstructs our desired signal, while suppressing the last two parts of the sum. For the ideal performance on dereverberation, interference suppression and noise reduction, the desire optimal frequency responses are to satisfy the following conditions

$$\begin{aligned} \sum_{m=1}^{M} W_m(\omega) H_{0,m}(\omega) &= G_D(\omega), \\ \sum_{m=1}^{M} W_m(\omega) H_{n,m}(\omega) &= 0,\ n = 1, \ldots, N-1, \\ \sum_{m=1}^{M} W_m(\omega) V_m(\omega) &= 0, \end{aligned} \tag{10}$$

where $G_D(\omega)$ denotes the direct path transfer function from the SOI point to the beamformer output reference point. In LCMV beamforming, the frequency responses $W_m(\omega)$ are adjusted based on the statistics of the sensor array measured signals, the expression to be minimized is the power of the background noise and the cost function can be defined as

$$\boldsymbol{E}(\boldsymbol{W}(\omega)) = ||\boldsymbol{W}^H(\omega)\boldsymbol{V}(\omega)||^2, \tag{11}$$

where $\boldsymbol{V}(\omega) = [V_1(\omega)\ V_2(\omega)\ \ldots\ V_M(\omega)]^T$ is the received noise vector. If we formulate the conditions on dereverberation and interference suppression as constraints, the frequency domain LCMV beamformer design problem can be established as

$$\begin{aligned} \min_{\boldsymbol{W}(\omega)\in\mathcal{C}^M} \quad & \boldsymbol{E}(\boldsymbol{W}(\omega)) \\ s.t. \quad & \boldsymbol{H}^H(\omega)\boldsymbol{W}(\omega) = \boldsymbol{G}(\omega), \end{aligned} \tag{12}$$

where the constraint matrix $\boldsymbol{H}(\omega)$ is constructed by the corresponding conditions on dereverberation and interference suppression in (10), defined as

$$\boldsymbol{H}(\omega) = \begin{pmatrix} H_{0,1}(\omega) & H_{1,1}(\omega) & \cdots & H_{N-1,1}(\omega) \\ H_{0,2}(\omega) & H_{1,2}(\omega) & \cdots & H_{N-1,2}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ H_{0,M}(\omega) & H_{1,M}(\omega) & \cdots & H_{N-1,M}(\omega) \end{pmatrix}, \tag{13}$$

and $\boldsymbol{G}(\omega) = [G_D(\omega)\ 0]^T$ is the response vector.

The problem (12) provides a general framework on the design of LCMV beamformer with respect to the frequency response functions. It is defined on the complex field with the complex variables $W_m(\omega),\ m = 1, \ldots, M$. To formulate the specific real field optimization model for the proposed LCMV beamformer design problem (12), we separate the complex transfer functions $H_{n,m}(\omega),\ n = 0, 1, \ldots, N-1$, noise components $V_m(\omega)$ and frequency response variables $W_m(\omega),\ m = 1, \ldots, M$ into real and imaginary parts. By denoting

$$\begin{aligned} H_{n,m}^{\mathrm{Re}}(\omega) &= \mathrm{real}(H_{n,m}(\omega)), & H_{n,m}^{\mathrm{Im}}(\omega) &= \mathrm{imag}(H_{n,m}(\omega)), \\ V_m^{\mathrm{Re}}(\omega) &= \mathrm{real}(V_m(\omega)), & V_m^{\mathrm{Im}}(\omega) &= \mathrm{imag}(V_m(\omega)), \\ W_m^{\mathrm{Re}}(\omega) &= \mathrm{real}(W_m(\omega)), & W_m^{\mathrm{Im}}(\omega) &= \mathrm{imag}(W_m(\omega)), \\ G_D^{\mathrm{Re}}(\omega) &= \mathrm{real}(G_D(\omega)), & G_D^{\mathrm{Im}}(\omega) &= \mathrm{imag}(G_D(\omega)), \end{aligned}$$

and stacking the real and imaginary parts $W_m^{\mathrm{Re}}(\omega)$ and $W_m^{\mathrm{Im}}(\omega)$ of frequency response into a new real variable vector $\chi_m(\omega) = (W_m^{\mathrm{Re}}(\omega)\ W_m^{\mathrm{Re}}(\omega))^T$, we can rewrite the beamforming

conditions (10) as

$$\begin{aligned}
&\sum_{m=1}^{M}\begin{pmatrix} H_{0,m}^{\mathrm{Re}}(\omega) & -H_{0,m}^{\mathrm{Im}}(\omega) \\ H_{0,m}^{\mathrm{Im}}(\omega) & H_{0,m}^{\mathrm{Re}}(\omega) \end{pmatrix}\chi_m(\omega) = \begin{pmatrix} G_D^{\mathrm{Re}}(\omega) \\ G_D^{\mathrm{Im}}(\omega) \end{pmatrix}, \\
&\sum_{m=1}^{M}\begin{pmatrix} H_{n,m}^{\mathrm{Re}}(\omega) & -H_{n,m}^{\mathrm{Im}}(\omega) \\ H_{n,m}^{\mathrm{Im}}(\omega) & H_{n,m}^{\mathrm{Re}}(\omega) \end{pmatrix}\chi_m(\omega) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad n = 1, \ldots, N-1, \\
&\sum_{m=1}^{M}\chi_m(\omega)^T\left(\begin{pmatrix} V_m^{\mathrm{Re}}(\omega) \\ -V_m^{\mathrm{Im}}(\omega) \end{pmatrix}\begin{pmatrix} V_m^{\mathrm{Re}}(\omega) \\ -V_m^{\mathrm{Im}}(\omega) \end{pmatrix}^T + \begin{pmatrix} V_m^{\mathrm{Im}}(\omega) \\ V_m^{\mathrm{Re}}(\omega) \end{pmatrix}\begin{pmatrix} V_m^{\mathrm{Im}}(\omega) \\ V_m^{\mathrm{Re}}(\omega) \end{pmatrix}^T\right)\chi_m(\omega) = 0.
\end{aligned} \tag{14}$$

By stacking all the matrices and vectors together, we establish the transformed LCMV beamformer design problem for solving optimal frequency responses as

$$\begin{aligned}
\min_{\boldsymbol{\chi}(\omega)\in\mathcal{R}^{2M}} \quad & \boldsymbol{\chi}(\omega)^T\boldsymbol{R}_{VV}(\omega)\boldsymbol{\chi}(\omega) \\
s.t. \quad & \bar{\boldsymbol{H}}^T(\omega)\boldsymbol{\chi}(\omega) = \bar{\boldsymbol{G}}(\omega),
\end{aligned} \tag{15}$$

where $\boldsymbol{\chi}(\omega)$, $\boldsymbol{R}_{VV}(\omega)$, $\bar{\boldsymbol{H}}(\omega)$ and $\bar{\boldsymbol{G}}(\omega)$ are the vectors and matrices corresponding to conditions (14). As a result, for each of the frequency ω, this problem can be tackled by the quadratic programming technique and obtain the optimal beamformer output $Y_{opt}(\omega)$ for a given $\boldsymbol{\lambda}$ by substituting $W_{m,opt}(\omega)$ into (4).

IV. Hybrid Descent Method

Based on the performance limit estimation of the LCMV beamformer, we have estimated the optimal beamformer output for measuring the fitness of the given $\boldsymbol{\lambda}$ irrespective of the filter length. The original configuration design problem (6) with respect to placement variables $\boldsymbol{\lambda}$ and filter coefficients $\boldsymbol{w}$ is converted firstly into a new optimization problem as

$$\begin{aligned}
\min_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda},\boldsymbol{W}(\omega)\in\mathbb{C}^M} \quad & \bar{\boldsymbol{F}}(\boldsymbol{\lambda}, \boldsymbol{W}(\omega)) \\
s.t. \quad & ||\boldsymbol{r}_i - \boldsymbol{r}_j||^2 \geq \varepsilon_0,\ i, j = 1, 2, \ldots, M,\ i \neq j,
\end{aligned} \tag{16}$$

where $\bar{\boldsymbol{F}}(\boldsymbol{\lambda}, \boldsymbol{W}(\omega))$ is the merit function (5) with the frequency responses determined by optimization problem (15). Considering in a bi-level manner, the optimal frequency responses $\boldsymbol{W}_{opt}(\omega)$ for (15) are fixed for a given $\boldsymbol{\lambda}$. Consequently the configuration design problem (16) is reduced to

$$\begin{aligned}
\max_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \quad & \mathcal{F}(\boldsymbol{\lambda}) = \bar{\boldsymbol{F}}(\boldsymbol{\lambda}, \boldsymbol{W}_{opt}(\omega)) \\
s.t. \quad & ||\boldsymbol{r}_i - \boldsymbol{r}_j||^2 \geq \varepsilon_0,\ i, j = 1, 2, \ldots, M,\ i \neq j.
\end{aligned} \tag{17}$$

The modified design problem (17) consists of the subproblem (15) for solving optimal frequency responses $\boldsymbol{W}_{opt}(\omega)$ and the main problem for solving the optimal placement $\boldsymbol{\lambda}_{opt}$. For the subproblem (15), the objective function is convex with respect to the variables $\boldsymbol{W}(\omega)$,

and the optimal solution can be solved effectively. Whereas, for the optimal placement, the merit function $\mathcal{F}(\boldsymbol{\lambda})$ has the placement variables $\boldsymbol{\lambda}$ nested inside the corresponding subproblem (15). It is a nonconvex optimization problem with respect to $\boldsymbol{\lambda}$ and cannot be tackled by gradient-based technique alone. Therefore, we introduce the genetic algorithm and propose a hybrid descent method as follows:

Algorithm 1.

1. *Given positive parameters $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, generate an initial placement $\boldsymbol{\lambda}^0$, and solve the subproblem (15) with $\boldsymbol{\lambda} = \boldsymbol{\lambda}^0$ to obtain the optimal frequency response $\boldsymbol{W}_{opt}(\omega)$ for the objective value as $\mathcal{F}(\boldsymbol{\lambda}^0)$. Set $k = 0$.*

2. *Take $\boldsymbol{\lambda}^k$ as one of the candidate point for the genetic algorithm, and execute Q iterations until a point $\boldsymbol{\lambda}^{\bar{k}}$ is obtained with the property $\mathcal{F}(\boldsymbol{\lambda}^{\bar{k}}) - \mathcal{F}(\boldsymbol{\lambda}^k) \leq -\varepsilon_1$.*

3. *Solve for the local minimum of $\mathcal{F}(\boldsymbol{\lambda})$ by using a gradient-based minimization method with $\boldsymbol{\lambda}^{\bar{k}}$ as the input point to get $\boldsymbol{\lambda}^{k+1}$, such that the objective function has a certain degree of decline, $\mathcal{F}(\boldsymbol{\lambda}^{k+1}) - \mathcal{F}(\boldsymbol{\lambda}^{\bar{k}}) \leq -\varepsilon_2$.*

4. *Set $k := k + 1$, return to Step* **2** *until convergence.*

In step **2** of Algorithm 1, the fitness function value for genetic algorithm is set to be the optimal value of the subproblem (15), and the genetic algorithm composes of five key steps as follows:

1. Population representation – The real-valued placement variables $\boldsymbol{\lambda}$ are initialized to construct the chromosomes and store an entire population in a single matrix with all chromosomes are of equal length.

2. Fitness assignment – The fitness values are derived from the optimal value of (16) via ranking or scaling.

3. Selection – Selection functions select a given number of placements from the current population, according to their fitness, and return a column vector to their indices.

4. Crossover – Crossover operators recombine pairs of individuals with given probability to produce offspring.

5. Mutation – Mutation operators apply random changes to individual parents to form offspring.

The overall procedure is outlined as follows:

Genetic algorithm

(a) Generate the initial placement chromosomes $\boldsymbol{\Lambda}^k$ via the last local optimal placement $\boldsymbol{\lambda}^k$,

evaluate (17) for all the individuals to obtain $\mathcal{F}(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \boldsymbol{\Lambda}^k$.

(b) Rank $\mathcal{F}(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \boldsymbol{\Lambda}^k$, if $\mathcal{F}(\boldsymbol{\lambda}^{\bar{k}}) - \mathcal{F}(\boldsymbol{\lambda}^k) \leq -\varepsilon_1$, stop; otherwise, go to next step (c).

(c) Select the individuals $\boldsymbol{\Lambda}_1^k$ by using certain selection functions to be parents.

(d) Combine two parents to form offspring $\boldsymbol{\Lambda}_2^k$ for the next generation by using the crossover operator.

(e) Apply random changes to individual parents to form offspring $\boldsymbol{\Lambda}_3^k$ by using mutation operator.

(e) Evaluate (17) for all of the new individuals $\boldsymbol{\Lambda}_3^k$; get $\mathcal{F}(\boldsymbol{\lambda}), \boldsymbol{\lambda} \in \boldsymbol{\Lambda}_3^k$ and return to (b).

V. Numerical examples

A. *Performance indicators*

To verify the effect of configuration designs, some indicators are needed to measure the performance of the beamformer outputs. There are many objective measures in the literature for evaluating the performance of speech enhancements [23]. Popular indicators include the segmented SNR (SNR_{seg}) [15], the PESQ score [30], the KurtR measure [36], [38], the NRR and the LLR measure [28]; these measures are demonstrated to be consistent in [38]. In our experiments, we simply use the SNR_{seg} the PESQ measure. In addition, we also define the generalized signal-to-noise, -interference and -reverberation ratio (SNIRR) for measuring the performance of speech quality degraded by reverberation, interference and background noise. And the segmental SNIRR measure can be defined similarly as

$$SNIRR_{seg} = \frac{10}{Q} \sum_{q=0}^{Q-1} \log_{10} \frac{||\mathbf{S}(q)||^2}{||\mathbf{S}(q) - \hat{\mathbf{S}}(q)||^2} \text{ dB}, \tag{18}$$

where $\mathbf{S}(q)$ and $\hat{\mathbf{S}}(q)$ represent the direct captured speech signal when the source signal is active alone without reverberation and the enhanced speech in the q-th time frame, respectively. The calculations are thresholded in interval $[-20\text{dB}, 35\text{dB}]$ to discard non-speech frames.

For the PESQ score, it is recommended by ITU-T for speech quality assessment of 3.2kHz handset telephony and narrow-band speech codecs [26], [27]. It is obtained by a linear combination of the average disturbance value D_{ind} and the average asymmetrical disturbance value A_{ind}

$$PESQ = 4.5 - 0.1 D_{ind} - 0.0309 A_{ind}. \tag{19}$$

The range of the PESQ score is 0.5 to 4.5, although for most cases the output range will be a MOS-like score, i.e., a score between 1.0 and 4.5.

The segmental SNIRR and the PESQ score are all pointed to the performance of speech quality. Specific indicators for measuring the performance on dereverberation, interference suppression and noise reduction can also be introduced similar to [37]. They are

$$Supp_{INT} = \frac{\int_{-\pi}^{\pi} P_{Y_I}(w)dw}{C_d \int_{-\pi}^{\pi} P_{S_I}(w)dw}, \quad Supp_{NOI} = \frac{\int_{-\pi}^{\pi} P_{Y_N}(w)dw}{C_d \int_{-\pi}^{\pi} P_{S_N}(w)dw}, \tag{20}$$

where the constant C_d is defined as

$$C_d = \frac{\int_{-\pi}^{\pi} P_{Y_S}(w)dw}{\int_{-\pi}^{\pi} P_{S_S}(w)dw},$$

and $P_{S_S}(w)$, $P_{S_I}(w)$ and $P_{S_N}(w)$ denote the spectral power estimate of the unfiltered signals when the source, interference and background noise active alone, respectively, while $P_{Y_S}(w)$, $P_{Y_I}(w)$ and $P_{Y_N}(w)$ denote the spectral power estimate of the filtered signals when the source, interference and background noise active alone, respectively. For dereverberation, we can define as

$$Supp_{REV} = \frac{\int_{-\pi}^{\pi} P_{Y_R}(w)dw}{C_d \int_{-\pi}^{\pi} P_{S_R}(w)dw}, \tag{21}$$

where $P_{S_R}(w)$ and $P_{Y_R}(w)$ denote the spectral powers estimate of the unfiltered and filtered reverberant signals active alone, respectively.

B. Acoustic environments

Assume the microphone array is deployed as an indoor product. A simple rectangular room with dimension $12m \times 6m \times 3m$ is defined for modelling. A fraction of sound wave are absorbed by the walls, floor and ceiling. These are the energy loss from the room and the fractional loss is characterized by several absorption coefficients, whereas, the overall effect can be represented by the reverberation time T_{60}. The SOI is imposed in $(0, 0, 1m)$, one interferer is placed at $(0, 1m, 1m)$, and the noise is located at $(0, -1m, 1m)$. The centre of the room is the origin. Note that the noise location is not used directly during beamformer design. An illustration of the room setup is depicted in Fig. 2. All the RIRs from the source points to sensor array are estimated by the fast-ISM simulator [18]. Three different kinds of room acoustics: $T_{60} = \{0.05s, 0.1s, 0.2s\}$ are studied. A male speech is employed as the SOI signal, while a female speech is used as the interference (INT). Both of them have $8000Hz$ sampling frequency. White noise is assumed to be the NOI.

We define a 9-elements microphone array settled in the plane $\Lambda = \{1m \leq x \leq 5m, -2.5m \leq y \leq 2.5m, z = 1m\}$ as the feasible region to capture the SOI, INT and NOI signals, and use one of the microphone as the reference point. Commonly deployed beamformer configurations, such as uniform linear array $\boldsymbol{\lambda}_{Lin}$, will be used for comparison.

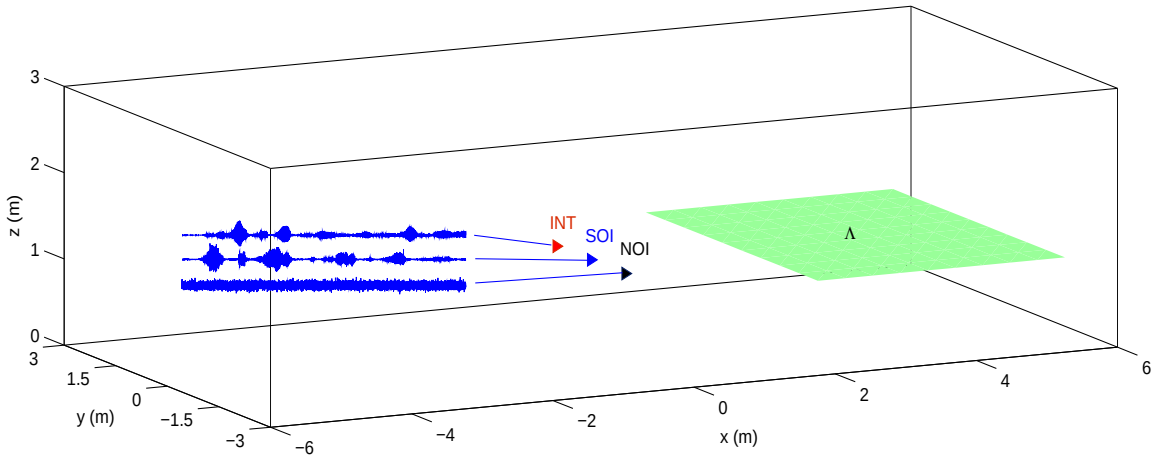


Fig. 2: Setup of the acoustic room system.

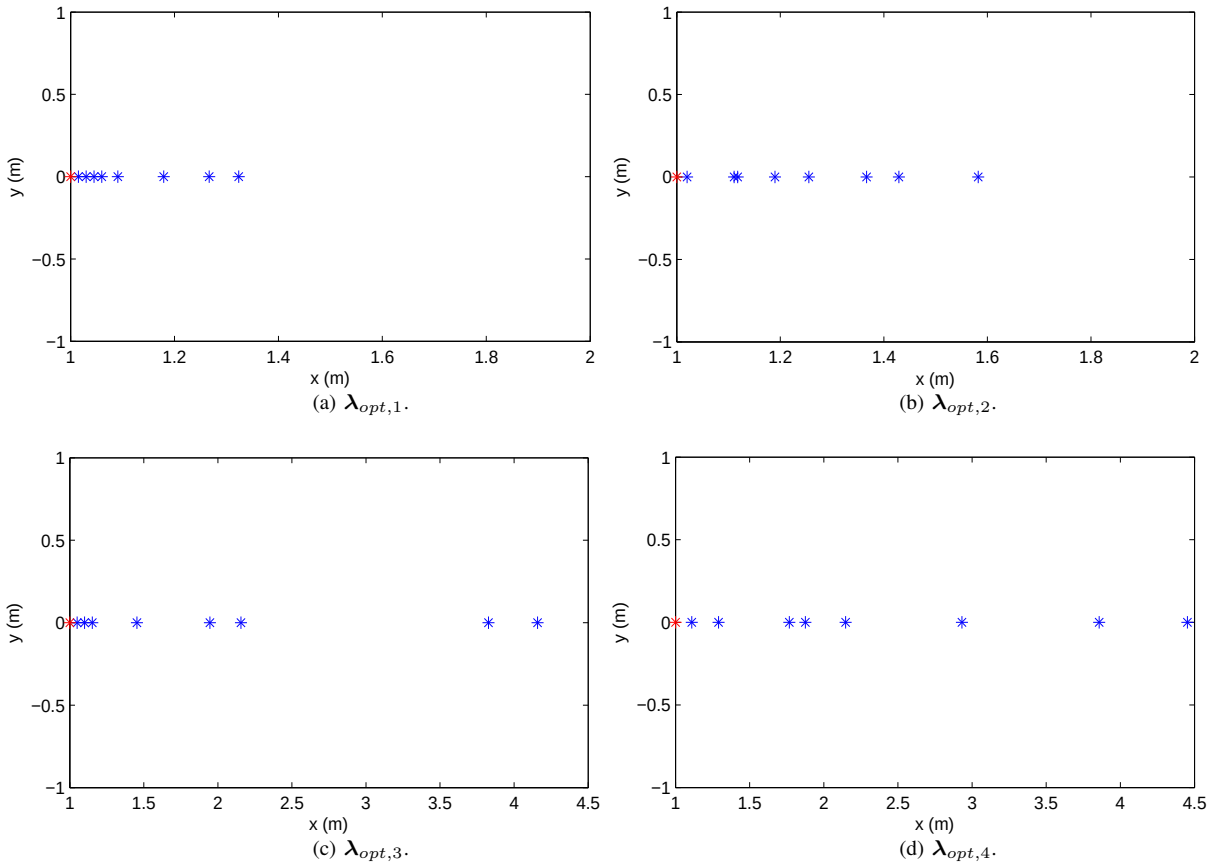
C. Optimal configurations

We first apply the proposed design method with $T_{60} = 0s$, which corresponds to the scenario of open area without reverberation. The optimal configuration $\boldsymbol{\lambda}_{opt,1}$ is shown in Fig. 3(a). We can see that all the microphone elements are converged to the vertical line in front of the source point. This result is in good agreement with [11]. Then, we employ this solution $\boldsymbol{\lambda}_{opt,1}$ as the initial points for solving problems in the reverberant environment. The converged solutions with reverberation time $T_{60} = \{0.05s, 0.1s, 0.2s\}$ are depicted in the Fig. 3(b)-(d). From the figures, we can see that all the converged placements are still arranged on the vertical line in front of the source points. Moreover, the configuration starts to spread out with reverberation. For $T_{60} = 0.2s$, one microphone is even moved close to the wall.

D. Beamformer performance

In this section, we compare the performances for different designs based on the indicators outlined above. We use the five indicators: reverberation suppression $Supp_{REV}$, interference suppression $Supp_{INT}$, noise suppression $Supp_{NOI}$, segmental SNIRR and PESQ scores to measure the overall performance. The overall results are listed in Table I. In the table, the best results of the enhancement are in boldface, while the other designs are for comparison. Moreover, all the designed beamformers have better performance on speech enhancement than the common uniform linear array $\boldsymbol{\lambda}_{Lin}$.

For illustration, we also plot the filtered results on the reverberation suppression, interference suppression, noise suppression and speech enhancement with $T_{60} = 0.2s$ in Fig. 4. From the results, it can be seen that the optimized beamformers can achieve effective suppressions on reverberation, interference and noise, and finally, enhance the SOI signal.



(a) $\boldsymbol{\lambda}_{opt,1}$. (b) $\boldsymbol{\lambda}_{opt,2}$.

(c) $\boldsymbol{\lambda}_{opt,3}$. (d) $\boldsymbol{\lambda}_{opt,4}$.

Fig. 3: Configurations of the converged designs.

VI. Conclusion

In this paper, we have studied the microphone array configuration design problem in a reverberant environment. We have employed the room impulse responses in the design to account for reverberation effect. Based on the LCMV beamforming technique, the configuration design optimization problem has been formulated. Moreover, to eliminate the filter length effect on the designs, the performance limit on filter length has been studied for LCMV beamformers and a design problem with only location variables has been proposed. A hybrid descent method with genetic algorithm has been proposed for solving this nonconvex problem. From the experimental results, we found that the optimal microphone elements are all arranged on the vertical line in front of the source points. Moreover, the configurations are expanded farther as reverberation increases and microphones are located closer to the walls. The proposed method can be applied directly for product designs deploying a set of microphones for indoor applications. As a future extension, it is of interest to study multiple noise sources in the optimization process.

TABLE I: Summary of the designed beamformer performances.

T_{60}	$\boldsymbol{\lambda}$	$Supp_{REV}$	$Supp_{INT}$	$Supp_{NOI}$	$SNIRR_{seg}$		$PESQ$	
					Unfiltered	*Filtered*	*Unfiltered*	*Filtered*
0	$\boldsymbol{\lambda}_{Lin}$	—	57.6870	17.3143	-4.0122	10.1284	1.3565	2.7864
	$\boldsymbol{\lambda}_{opt,1}$	**—**	**64.5225**	**55.2305**	**-4.0533**	**31.5169**	**1.3575**	**4.4891**
	$\boldsymbol{\lambda}_{opt,2}$	—	52.3383	49.1290	-4.0533	28.9000	1.3575	4.4579
	$\boldsymbol{\lambda}_{opt,3}$	—	33.2180	35.8123	-4.0526	20.6136	1.3575	3.9036
	$\boldsymbol{\lambda}_{opt,4}$	—	32.5225	34.3196	-4.0526	20.0587	1.3575	3.9224
0.05	$\boldsymbol{\lambda}_{Lin}$	44.1851	44.2679	15.5927	-4.0912	7.7772	1.3529	2.6682
	$\boldsymbol{\lambda}_{opt,1}$	33.8200	46.7055	47.0113	-4.1269	24.8553	1.3527	4.4356
	$\boldsymbol{\lambda}_{opt,2}$	**42.2344**	**50.7277**	**46.2392**	**-4.1269**	**27.3810**	**1.3527**	**4.4034**
	$\boldsymbol{\lambda}_{opt,3}$	33.5334	32.2094	35.3161	-4.1262	20.1291	1.3527	3.7127
	$\boldsymbol{\lambda}_{opt,4}$	29.5846	30.2106	34.7869	-4.1262	19.1560	1.3527	3.6960
0.1	$\boldsymbol{\lambda}_{Lin}$	19.1569	27.9448	14.1948	-4.6129	5.7986	1.3715	2.5198
	$\boldsymbol{\lambda}_{opt,1}$	10.4564	24.3906	27.6087	-4.6348	10.1322	1.3765	3.0990
	$\boldsymbol{\lambda}_{opt,2}$	16.4976	25.4371	26.8072	-4.6427	11.7183	1.3760	3.1170
	$\boldsymbol{\lambda}_{opt,3}$	**21.3615**	**29.0429**	**28.8280**	**-4.6310**	**14.9132**	**1.3740**	**3.1248**
	$\boldsymbol{\lambda}_{opt,4}$	21.0103	27.7977	28.0733	-4.6313	14.0199	1.3746	3.1207
0.2	$\boldsymbol{\lambda}_{Lin}$	14.8547	21.3085	13.0079	-5.1495	3.9352	1.3699	2.3513
	$\boldsymbol{\lambda}_{opt,1}$	15.3812	20.6743	24.1998	-5.0849	8.9280	1.3739	2.8609
	$\boldsymbol{\lambda}_{opt,2}$	16.8513	20.7430	23.9908	-5.0849	9.0894	1.3739	2.8857
	$\boldsymbol{\lambda}_{opt,3}$	20.0668	24.3083	24.6788	-5.0844	10.6398	1.3740	2.8854
	$\boldsymbol{\lambda}_{opt,4}$	**20.4096**	**25.2930**	**24.8435**	**-5.0845**	**11.1968**	**1.3740**	**2.9366**

ACKNOWLEDGEMENT

This research is supported by RGC Grant PolyU. (152200/14E) and the research committee of the Hong Kong Polytechnic University.

REFERENCES

[1] J. Allen and D. Berkley. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Amer.*, 65(4):943–950, 1979.

[2] D. Ayllón, R. Gil-Pita, M. Utrilla-Manso, and M. Rose-Zurere. An evolutionary algorithm to optimize the microphone array configuration for speech acquisition in vehicles. *Eng. Appl. Artif. Intel.*, 34:37–44, 2014.

[3] J. Benesty, J. Chen, and Y. Huang. *Microphone Array Signal Processing*. Springer Verlag, Berlin, 2008.

[4] J. Borish. An extension of the image model to arbitrary polyhedra. *J. Acoust. Soc. Amer.*, 75:1827–1836, 1984.

[5] J. Capon. High resolution frequency-wavenumber spectrum analysis. *Proc. IEEE*, 57(8):1408–1418, 1969.

[6] K.S. Chen, X.H. Yun, Z.S. He, and C.L. Han. Synthesis of sparse planar arrays using modified real genetic algorithm. *IEEE T. Antenn. Propag.*, 55(4):1067–1073, 2007.

[7] M.F. Delgado, J.R. Gonzalez, R. Iglesias, S. Barro, and F.A. Pena. Fast array thinning using global optimization methods. *J. Electromagn. Wave. Appl.*, 24(16):2259–2271, 2010.

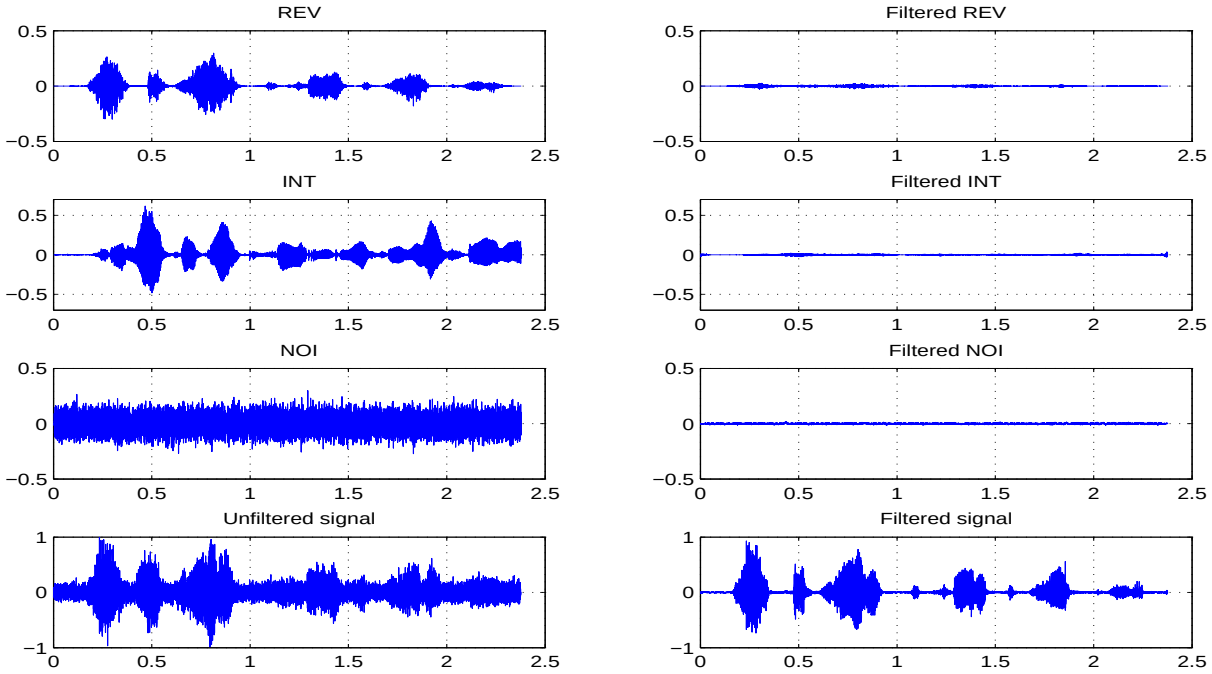


Fig. 4: Limit performance of beamformer in optimal placement at $T_{60} = 0.2$s.

[8] G. Doblinger. Optimized design of interpolated array and sparse array wideband beamformers. In *16th European Signal Processing Conference (EUSIPCO 2008)*, Lausanne, Switzerland, 2008.

[9] M. Er and A. Cantoni. Derivative constraints for broad-band element space antenna array processors. *IEEE Trans. Acoust., Speech, Signal Process.*, 31(6):1378–1393, 1983.

[10] Z.G. Feng, K.F.C. Yiu, and S. Nordholm. A two-stage method for the design of near-field broadband beamformer. *IEEE Trans. Signal Process.*, 59(8):3647–3656, 2011.

[11] Z.G. Feng, K.F.C. Yiu, and S. Nordholm. Placement design of microphone arrays in near-field broadband beamformers. *IEEE Trans. Signal Process.*, 60(3):1195–1204, 2012.

[12] Z.G. Feng, K.F.C. Yiu, and S. Nordholm. Performance limit of broadband beamformer designs in space and frequency. *J. Optimiz. Theory App.*, 164(1):316–341, 2015.

[13] O. Frost. An algorithm for linearly constrained adaptive array processing. *Proc. IEEE*, 60(8):926–935, 1972.

[14] S. Gannot and I. Cohen. Adaptive beamforming and postfiltering. In J. Benesty, M.M. Sondhi, and Y. Huang, editors, *Springer Handbook of Speech Processing*, chapter 47. Berlin, Germany: Springer-Verlag, 2008.

[15] J. Hansen and B. Pellom. An effective quality evaluation protocol for speech enhancement algorithms. In *Proc. Internat. Conf. on Spoken Language Processing*, pages 2819–2822, San Francisco, CA, 1998.

[16] R.L. Haupt. Thinned arrays using genetic algorithms. *IEEE Trans. Antenn. Propag.*, 42:993–999, 1994.

[17] C. Kumar, S. Rao, and A. Hoorfar. Optimization of thinned phased arrays using evolutionary programming. In *Evolutionary Programming VII*, volume 1447 of *Lecture Notes in Computer Science*, pages 157–166. Springer Berlin / Heidelberg, 1998.

[18] E.A. Lehmann and A.M. Johansson. Diffuse reverberation model for efficient image-source simulation of room impulse responses. *Trans. Audio, Speech and Lang. Proc.*, 18:1429–1439, 2010.

[19] H. Lehnert and J. Blauert. Principles of binaural room simulation. *Appl. Acoust.*, 36(3-4):259–291, 1992.

[20] Z.B. Li and K.F.C. Yiu. A least-squares indoor beamformer design. *Pacific J Optim*, 9(4):697–707, 2013.

[21] Z.B. Li, K.F.C. Yiu, and Z.G. Feng. A hybrid descent method with genetic algorithm for microphone array placement design. *Applied Soft Computing*, 13(3):1486–1490, 2013.

[22] Z.B. Li, K.F.C. Yiu, and S. Nordholm. On the indoor beamformer design with reverberation. *IEEE Trans Audio, Speech Lang Proc*, 22(8):1225–1235, 2014.

[23] P. Loizou. *Speech Enhancement: Theory and Practice*. CRC Press, Boca Raton, FL, 2007.

[24] J.T. Mayhan. Thinned array configurations for use with satellitebased adaptive antennas. *IEEE Trans. Antenn. Propag.*, 28(6):846–856, 1980.

[25] G. Oliveri, M. Donelli, and A. Massa. Linear array thinning exploiting almost difference sets. *IEEE Trans. Antenn. Propag.*, 57(12):3800–3812, 2009.

[26] ITU-T P.835. *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*. ITU-T Recommendation P.835, 2003.

[27] ITU-T P.862. *Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs*. ITU-T Recommendation P.862, 2000.

[28] S. Quackenbush, T. Barnwell, and M. Clements. *Objective Measures of Speech Quality*. Prentice Hall, Englewood Cliffs, NJ, 1988.

[29] A. Razavi and K. Forooraghi. Progress in electromagnetics research. *IEEE Trans. Antenn. Propag.*, 78:61–71, 2008.

[30] A. Rix, J. Beerends, M. Hollier, and A. Hekstra. Perceptual evaluation of speech quality (pesq), a new method for speech quality assessment of telephone networks and codecs. In *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'01)*, volume 2, pages 749–752, Barcelona, Spain, 2001.

[31] P. Rocca and R. Haupt. Dynamic array thinning for adaptive interference cancellation. In *The Fourth European Conference on Antennas and Propagation (EuCAP2010)*, pages 1–3, Barcelona, Spain, 2010.

[32] A. Sarajedini. Adaptive array thinning for space-time beamforming. In *The Thirty-Third Asilomar Conference on Signals, Systems, and Computers*, pages 1572–1576, Pacific Grove, CA, USA, 1999.

[33] U.P. Svensson. Modelling acoustic spaces for audio virtual reality. In *Workshop on Model based Processing and Coding of Audio (MPCA)*, pages 109–116, Leuven, Belgium, 2002.

[34] A. Trucco. Weighting and thinning wide-band arrays by simulated annealing. *Ultrasonics*, 40:485–489, 2002.

[35] A. Trucco and V. Murino. Stochastic optimization of linear sparse arrays. *IEEE J. Oceanic Eng.*, 24(3):291–299, 1999.

[36] Y. Uemura, Y. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo. Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics. In *Proc. Internat. Workshop on Acoustic Echo and Noise Control (IWAENC'08)*, Seattle, USA, 2008.

[37] K.F.C. Yiu, N. Grbic, K.L. Teo, and S. Nordholm. A new design method for broadband microphone arrays for speech input in automobiles. *IEEE Signal Proc. Let.*, 9(7):222–224, 2002.

[38] P.C. Yong, S. Nordholm, and H.H. Dam. Optimization and evaluation of sigmoid function with *a priori* SNR estimate for real-time speech enhancement. *Speech Commun.*, 55(2):358–376, 2013.