

On the indoor beamformer design with reverberation

Zhibao Li , Ka Fai Cedric Yiu and Sven Nordholm

Abstract

Beamforming remains to be an important technique for signal enhancement. For applications in open space, the transfer function describing waves propagation has an explicit expression, which can be employed for beamformer design. However, the function becomes very complex in an indoor environment due to the effects of reverberation. In this paper, this problem is discussed. A method based on the image source method (ISM) is applied to model the room impulse responses (RIRs), which will act as the transfer function between source and sensor. The indoor beamformer design problem is formulated as a minimax optimization problem. We propose and study several optimization models based on the L_1 -norm to design the beamformer. We found that it is advantageous to separate early and late reverberations in the design process and better designs can be achieved. Several numerical experiments are presented using both simulated data and real recordings to evaluate the proposed methods.

Key words: Beamformer design, signal enhancement, reverberation, dereverberation.

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This paper is supported by RGC Grant PolyU. (5301/12E) and the research committee of the Hong Kong Polytechnic University. The authors wish to thank Dr. Anders Johansson, Dr. Siow Yong Low and Dr. Nedelko Grbić for their helps in sound recordings. The authors also would like to thank the anonymous reviewers for their useful comments and suggestions.

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; tel: +852-34008981; fax: +852-23629045; e-mail: macyiu@polyu.edu.hk

Sven Nordholm is with the Department of Electrical and Computer Engineering, Curtin University, Perth, Australia; tel: +618 92667439; e-mail: S.Nordholm@curtin.edu.au

I. INTRODUCTION

Beamforming provides a versatile form of spatial filtering by an array of sensors. Beamformers are often deployed in applications such as teleconferencing, hands-free communications, speech recognition and hearing aids. There are numerous algorithms dedicated to the design of beamformers in the literature. Many of the early beamforming techniques were developed for a spectrum of applications including wireless communications [1]–[3]. Wave propagations from the source point to the sensors are described by simple but elegant functions. When it comes to indoor acoustic applications, this creates a problem in the accuracy of the designs. Indoor sound propagation is a very complex phenomenon in the enclosure where the sound conducting medium is bounded on all sides by walls, ceiling and floor. Performance requirements on the beamformers are likely fail if a simple transfer function is employed in the design process. In recent years, many studies have been carried out on the speech dereverberation, noise reduction and source localization in reverberant environments [4], [5]. However, applying those techniques in a multi-microphone setting is still a challenging task.

In the design of microphone array for speech acquisition in reverberant environments, a mixed near-field/far-field technique is developed to solve the beamformer design problem [6], in which a near-field beamformer was designed to pass on a desired signal from a chosen near-field source location while a far-field beamformer is employed to suppress room reverberation. However, the study was based on the assumption that most interference/reverberation are generated from the far-field and the RIRs calculated are generated by Legendre function and spherical Hankel function. To improve the array performance in reverberant enclosures, a multiple beamforming and matched-filtering technique was proposed in [7], and the performance of the matched-filter array (MFA) processing in real rooms was investigated in [8]. In the MFA approach, the output of each microphone is processed by a time inverse of the impulse response with fixed delaying truncation from the focal point to the microphone, and the array output is the summation of outputs from each matched-filter, where the RIRs are measured by using the maximal length sequences (MLS) [9] as excitatory signals. These studies provide good insights for dereverberation, but have limitations such as inaccurate estimation of RIRs. Another common idea to deal with this problem is to design beamformers based on approximating the multiplicative transfer function (MTF) [10], [11], and more recently, the relative transfer function (RTF) estimator by applying the method proposed in [12]. In the past decades, various methods have been developed to identify the RTFs; more details can be found in [13] and the references therein.

Indeed, the sound field can be described by wave equation and appropriate boundary conditions for

the walls. However, this wave-based acoustical modelling approach is computationally demanding and is particularly useful for simulating low frequency sound fields [14]–[16]. Another way of modelling room acoustics is geometrical acoustics modelling techniques, such as ray tracing and the image-source method (ISM) [17]–[20]. They have been widely used to model the acoustic field of enclosed spaces. The corresponding predictions are valid for frequencies above the Schroeder’s frequency [21]–[23], $f_{Sch} = 2000\sqrt{T_{60}/V}(Hz)$ (T_{60} : reverberation time in second, V : volume in m^3). For instance, in an $8m \times 4m \times 3m$ room with reverberation time $T_{60} = 0.3s$, the Schroeder’s frequency is about $f_{Sch} \approx 112Hz$. In audible frequencies, the frequency range of $300Hz$ to $3400Hz$ has been found to be the most important for speech intelligibility and speech recognition [24], which means the geometrical acoustics modelling techniques are valid for the frequencies of interest in such rooms. The ISM based modelling originally proposed for rectangular enclosures by J. Allen and D. Berkley in 1979 [19] is viewed as a simple but efficient approach for simulating room acoustics. However, the required simulation times grow exponentially with the reflection order, and the associated computational costs is a drawback of the original image-source implementation. E.A. Lehmann and A.M. Johansson [25] proposed a diffuse reverberation model (fast-ISM) recently to reduce the computational cost by modelling the reverberation tail as decaying random noise, while using the ISM simulator for the computation of the early reflections.

In this paper, we study the indoor beamformer design problem, where the room impulse responses (RIRs) are estimated by an efficient room simulator developed in [25]. We transform the indoor beamformer design problem into a minimax optimization problem, where the maximum of the cost is chosen over the constrained class of position and frequency, while the minimum is taken over the set of filter coefficients. The minimax filter design problem is equivalent to a semi-infinite linear programming problem, so that we can convert it into a constrained linear programming problem by using the discretization technique. Since the propagation can be divided into 3 regimes, namely the direct path, early reflections and late reverberation, we propose several optimization models based on the L_1 -norm to design indoor beamformers by considering different combinations of the desired response and the reverberation. We show that the optimization model based on the separation of the propagation path into all 3 regimes performs better than the other existing models. Moreover, we demonstrate that early and late reverberation behave in a Pareto manner by numerical calculations, in the sense that there is no optimized solution that suppresses both effectively based on experimental results. It essentially becomes a multi-criteria problem and there is a trade-off in suppressing more on either one. As a result, we propose a multi-criteria optimization method which is able to obtain all the Pareto optima.

The rest of the paper is organized as follows. In Section II, the indoor beamformer design problem

is formulated as a minimax optimization problem by using the RIRs, where the RIRs are estimated by the efficient ISM based room simulator. In Section III, we construct the linear programming problem to implement the numerical model for the design of indoor beamformers based on the L_1 -norm. Two separation models based on a RIR separation scheme are proposed to improve beamforming performances. In Section IV, we present some numerical experiments to show that the proposed methods are effective, and carry out comparative studies among these methods. Finally, conclusions and summary of this paper are given in Section V.

II. PROBLEM FORMULATION

Assume a microphone array containing N elements with L -tap FIR filter behind each element is operating in a small rectangular room. Let the microphone elements be at the fixed points as $\mathbf{r}_i = (x_i, y_i, z_i)$, $i = 1, 2, \dots, N$. If the signals received by the microphone array are sampled synchronously at the rate of f_s per second, the frequency responses of these FIR filters may be defined as

$$H_i(\mathbf{h}, f) = \mathbf{h}_i^T \mathbf{d}_0(f), \quad i = 1, 2, \dots, N, \quad (1)$$

where \mathbf{h}_i is a vector consisting of the i -th L -tap FIR filter

$$\mathbf{h}_i = [h_i(0), h_i(1), \dots, h_i(L-1)]^T,$$

and $\mathbf{d}_0(f)$ is the vector defined as

$$\mathbf{d}_0(f) = [1, e^{-j\frac{2\pi f}{f_s}}, \dots, e^{-j\frac{2\pi f}{f_s}(L-1)}]^T.$$

The beamformer output can be obtained by using the room impulse responses (RIRs) and FIR-filter responses $H_i(\mathbf{h}, f)$. Denote the frequency domain RIR of the i -th microphone as $R_i(\mathbf{r}, f)$, where $\mathbf{r} = (x, y, z)$ is a space point, suppose the desired response is $G_d(\mathbf{r}, f)$, then our target is to find a group of coefficients $\mathbf{h} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N]^T$ for the FIR filters, such that the beamformer output

$$G(\mathbf{r}, f) = \sum_{i=1}^N H_i(\mathbf{h}, f) R_i(\mathbf{r}, f) = \mathbf{R}^T(\mathbf{r}, f) \mathbf{H}(\mathbf{h}, f),$$

is close to the desired response $G_d(\mathbf{r}, f)$, where $\mathbf{R}^T(\mathbf{r}, f) = [R_1(\mathbf{r}, f), \dots, R_N(\mathbf{r}, f)]^T$ is the frequency domain RIR vector, and $\mathbf{H}(\mathbf{h}, f) = [H_1(\mathbf{h}, f), \dots, H_N(\mathbf{h}, f)]^T$ is the frequency domain filter response vector defined in (1).

Selection of good criteria to measure the effectiveness of the designed beamformer is important. A simple model for the indoor beamformer design problem can be described as the following optimization

problem

$$\min_{\mathbf{h} \in \mathbb{R}^{N \times L}} \max_{(\mathbf{r}, f) \in \Omega} \|\mathbf{R}^T(\mathbf{r}, f)\mathbf{H}(\mathbf{h}, f) - G_d(\mathbf{r}, f)\|_p^p, \quad (2)$$

where $p > 0$, Ω is the given space-frequency domain for the definition of desired response function $G_d(\mathbf{r}, f)$. The problem formulation (2) is a general form to describe the beamformer design problem, it is to find the optimal filter coefficients \mathbf{h} to minimise the maximum error between the beamformer response and desired response over the considered region Ω .

When the desired response $G_d(\mathbf{r}, f)$ is given, then the major task is to estimate $\mathbf{R}(\mathbf{r}, f)$ for all $(\mathbf{r}, f) \in \Omega$. The following direct path transfer function

$$T(\mathbf{r}, f) = \frac{1}{\|\mathbf{r} - \mathbf{r}_i\|} e^{\frac{-j2\pi f \|\mathbf{r} - \mathbf{r}_i\|}{c}}, \quad (3)$$

is often deployed, where c is the speed of sound, which is the Green's function for the Helmholtz equation to describe the sound wave propagation in the acoustic free field [26]. However, if this design is deployed in a typical reverberant room, the beamforming performance deteriorates significantly when T_{60} increases. It is necessary to employ the corresponding information to design the indoor beamformer, and the estimation of RIRs is one of the key factor. However, the RIRs include lots of the reflections and reverberation due to the room acoustics. A typical example of RIR estimated from fast-ISM based room simulator is plotted in Fig. 1, showing a long tail of reflections and reverberation.

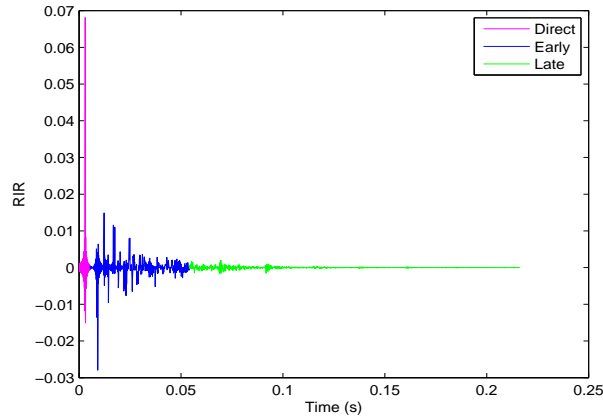


Fig. 1: Example of an estimated RIR.

It is noted that the estimated RIR can be separated into three parts, namely the direct path response, early reflections and late reverberation, with different delays, denoted by $\mathbf{R}_D(\mathbf{r}, f)$, $\mathbf{R}_E(\mathbf{r}, f)$ and $\mathbf{R}_L(\mathbf{r}, f)$, respectively. Therefore,

$$\mathbf{R}(\mathbf{r}, f) = \mathbf{R}_D(\mathbf{r}, f) + \mathbf{R}_E(\mathbf{r}, f) + \mathbf{R}_L(\mathbf{r}, f). \quad (4)$$

The direct path response $\mathbf{R}_D(\mathbf{r}, f)$ is defined to describe the wave propagation from the source point to the microphone element, and it can be determined by using the distance information. The early reflections $\mathbf{R}_E(\mathbf{r}, f)$ is used to describe the major energy decay of sound wave reflections to some extent, while the late reverberation $\mathbf{R}_L(\mathbf{r}, f)$ is used to model the diffuse reverberation decay. The RIR model requires the definition of a specific transition time (the cut-off point is denoted by t_c) between early reflections and late reverberation. The decomposition of RIR into early reflections and late reverberation is described in [25], [27]. According to the transition point t_c defined in [25], it is the time for which the overall acoustic energy in the RIR has decreased by a certain amount Δ_c (in dB)

$$t_c = E^{-1}(-\Delta_c)$$

where $E^{-1}(\tau)$ corresponds to the time lag τ such that $E(t) = \tau$. It effectively defines a room-dependent cut-off parameter that can be tuned to capture most of the early reflections in the specific environment under consideration.

If we model the desired response with the direct path response, then the early reflections and late reverberation are also expected to be suppressed to a certain extent. However, the suppression level cannot be controlled directly because reverberation is not a direct objective in the optimization process. In view of this, we can extract the reverberation part from the RIRs to modify the indoor beamformer design model (2) for desired response formulation and dereverberation respectively. One simple multi-criteria model is

$$\begin{aligned} \min_{\mathbf{h} \in \mathbb{R}^{N \times L}} \max_{(\mathbf{r}, f) \in \Omega} & \rho_1 \|\mathbf{R}_D^T(\mathbf{r}, f) \mathbf{H}(\mathbf{h}, f) - G_d(\mathbf{r}, f)\|_p^p \\ & + \rho_2 \|(\mathbf{R}_E(\mathbf{r}, f) + \mathbf{R}_L(\mathbf{r}, f))^T \mathbf{H}(\mathbf{h}, f)\|_p^p, \end{aligned} \quad (5)$$

where $\rho_1 > 0$ and $\rho_2 > 0$ are weighting parameters. They can be adjusted gradually within a range [28] to change the influence of the different objectives during the optimal beamformer design and obtain a set of solutions which constitutes the Pareto optima set. Since the direct path response model is fairly accurate, the separation of the direct path response beamforming and reflection parts suppression can reduce the influence of the measurement errors on the reflection parts of RIRs estimation, and the designed beamformer will have better robustness. In this model, the total reverberation is suppressed as a whole and is introduced directly into the objective function to ensure reverberation is reduced in the optimization process.

On the other hand, since the early reflections have most of the energy of the reverberation, a better

multi-criteria model for indoor beamformer design can be established as

$$\begin{aligned} \min_{\mathbf{h} \in \mathbb{R}^{N \times L}} \max_{(\mathbf{r}, f) \in \Omega} \rho_1 & \left\| \mathbf{R}_D^T(\mathbf{r}, f) \mathbf{H}(\mathbf{h}, f) - G_d(\mathbf{r}, f) \right\|_p^p \\ & + \rho_2 \left\| \mathbf{R}_E^T(\mathbf{r}, f) \mathbf{H}(\mathbf{h}, f) \right\|_p^p + \rho_3 \left\| \mathbf{R}_L^T(\mathbf{r}, f) \mathbf{H}(\mathbf{h}, f) \right\|_p^p, \end{aligned} \quad (6)$$

where $\rho_1 > 0$, $\rho_2 > 0$ and $\rho_3 > 0$ are weighting parameters. This is the most elaborate model to control early and late reverberation separately. Indeed, depending on applications, it is advantageous to have direct controls over both early and late reverberations, whereas they may yield different acoustic effects.

Since the L_1 -norm based model is effective in beamformer design problem [29], [30], it is introduced to measure the residuals of the desired response approximation and reverberation suppression for our indoor beamformer design based on the above developed models (2), (5) and (6), respectively.

III. THE IMPLEMENTATION MODELS

A. L_1 -norm implementation

It is noted that all the minimax problems (2), (5) and (6) are formulated in the continuous space-frequency region Ω , thus they are semi-infinite problems. And the general numerical schemes in dealing with them are the discretization methods and the reduction based methods [31]–[34]. For the discretization methods, in order to determine suitable multi-dimensional grids to solve the semi-infinite problems, sequences of adaptive meshes can be applied so that the meshes are refined gradually.

Let us consider first problem (2) with a multi-dimensional grid region Ω_M for approximating Ω with a uniform grid containing M mesh points in each dimension of the space-frequency domain, and with the frequency domain RIR vectors $\mathbf{R}(\mathbf{r}, f), \forall (\mathbf{r}, f) \in \Omega_M$ have been estimated by the fast-ISM room simulator. By combining the RIR vectors $\mathbf{R}(\mathbf{r}, f)$ with the vector \mathbf{d}_0 , we rearrange the expression as

$$\mathbf{R}^T(\mathbf{r}, f) \mathbf{H}(\mathbf{h}, f) = \mathbf{h}^T \mathbf{d}(\mathbf{r}, f). \quad (7)$$

Expanding the complex functions as

$$\begin{aligned} \mathbf{d}(\mathbf{r}, f) &= \mathbf{d}_1(\mathbf{r}, f) + j\mathbf{d}_2(\mathbf{r}, f), \\ G_d(\mathbf{r}, f) &= G_{d_1}(\mathbf{r}, f, L) + jG_{d_2}(\mathbf{r}, f, L), \end{aligned} \quad (8)$$

where $\mathbf{d}_1(\mathbf{r}, f)$, $\mathbf{d}_2(\mathbf{r}, f)$, $G_{d_1}(\mathbf{r}, f, L)$ and $G_{d_2}(\mathbf{r}, f, L)$ are the real and imaginary parts of $\mathbf{d}(\mathbf{r}, f)$ and $G_d(\mathbf{r}, f)$, respectively, and so on. Thus, denoting

$$\begin{aligned} \mathbf{u}(\mathbf{r}, f) &= \mathbf{h}^T \mathbf{d}_1(\mathbf{r}, f) - G_{d_1}(\mathbf{r}, f), \\ \mathbf{v}(\mathbf{r}, f) &= \mathbf{h}^T \mathbf{d}_2(\mathbf{r}, f) - G_{d_2}(\mathbf{r}, f). \end{aligned}$$

the minimax problem (2) for the beamformer design can be rewritten as

$$\min_{\mathbf{h} \in \mathbb{R}^{N \times L}} \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{u}(\mathbf{r}, f) + j\mathbf{v}(\mathbf{r}, f)|.$$

By using the L_1 -norm, the beamformer design problem can be expressed as

$$\min_{\mathbf{h} \in \mathbb{R}^{N \times L}} \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{u}(\mathbf{r}, f)| + |\mathbf{v}(\mathbf{r}, f)|. \quad (9)$$

To convert it into a linear programming problem, one approach is to introduce two new variables for controlling the real part and imaginary part separately as

$$\xi_{I,1} = \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{u}(\mathbf{r}, f)|, \quad \xi_{I,2} = \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{v}(\mathbf{r}, f)|,$$

then an implementation model for problem (9) can be established.

Model I

$$\begin{aligned} \min_{\mathbf{h} \in \mathbb{R}^{N \times L}, \xi_I \in \mathbb{R}^2} \quad & \xi_{I,1} + \xi_{I,2} \\ \text{s.t.} \quad & |\mathbf{u}(\mathbf{r}, f)| \leq \xi_{I,1}, \quad \forall (\mathbf{r}, f) \in \Omega_M, \\ & |\mathbf{v}(\mathbf{r}, f)| \leq \xi_{I,2}, \quad \forall (\mathbf{r}, f) \in \Omega_M. \end{aligned} \quad (10)$$

It can be formulated as the following standard linear programming model in matrix notation:

$$\begin{aligned} \min_{\mathbf{Z}_I \in \mathbb{R}^{N \times L+2}} \quad & \mathbf{c}^T \mathbf{Z}_I \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{r}, f) \mathbf{Z}_I - b \leq 0, \quad \forall (\mathbf{r}, f) \in \Omega_M, \end{aligned} \quad (11)$$

where

$$\mathcal{A}(\mathbf{r}, f) = \begin{pmatrix} \mathbf{d}_1(\mathbf{r}, f) & -1 & 0 \\ -\mathbf{d}_1(\mathbf{r}, f) & -1 & 0 \\ \mathbf{d}_2(\mathbf{r}, f) & 0 & -1 \\ -\mathbf{d}_2(\mathbf{r}, f) & 0 & -1 \end{pmatrix},$$

$$\mathbf{Z}_I = \begin{pmatrix} \mathbf{h} \\ \xi_{I,1} \\ \xi_{I,2} \end{pmatrix}, \mathbf{c} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} G_{d_1}(\mathbf{r}, f) \\ -G_{d_1}(\mathbf{r}, f) \\ G_{d_2}(\mathbf{r}, f) \\ -G_{d_2}(\mathbf{r}, f) \end{pmatrix}.$$

This is the current approach in the literature to achieve a target response without giving special consideration to reverberation. Therefore, only the direct path impulse response is needed in this formulation.

B. Multi-criteria models

Problems (5) and (6) can be formulated into linear programming problems like Model I based on the L_1 -norm measure. In the reverberation responses (4), denote

$$\mathbf{R}_R(\mathbf{r}, f) = \mathbf{R}_E(\mathbf{r}, f) + \mathbf{R}_L(\mathbf{r}, f).$$

Similar to (7) and (8), we rearrange

$$\begin{aligned} \mathbf{R}_D^T(\mathbf{r}, f)\mathbf{H}(\mathbf{h}, f) &= \mathbf{h}^T \mathbf{d}_D(\mathbf{r}, f), \\ \mathbf{R}_R^T(\mathbf{r}, f)\mathbf{H}(\mathbf{h}, f) &= \mathbf{h}^T \mathbf{d}_R(\mathbf{r}, f), \end{aligned}$$

and expand the complex functions as

$$\begin{aligned} \mathbf{d}_D(\mathbf{r}, f) &= \mathbf{d}_{D_1}(\mathbf{r}, f) + j\mathbf{d}_{D_2}(\mathbf{r}, f), \\ \mathbf{d}_R(\mathbf{r}, f) &= \mathbf{d}_{R_1}(\mathbf{r}, f) + j\mathbf{d}_{R_2}(\mathbf{r}, f). \end{aligned}$$

Similar to (9), we have

$$\begin{aligned} \mathbf{u}_D(\mathbf{r}, f) &= (\mathbf{h}^T \mathbf{d}_{D_1}(\mathbf{r}, f) - G_{d_1}(\mathbf{r}, f)), \\ \mathbf{v}_D(\mathbf{r}, f) &= (\mathbf{h}^T \mathbf{d}_{D_2}(\mathbf{r}, f) - G_{d_2}(\mathbf{r}, f)), \\ \mathbf{u}_R(\mathbf{r}, f) &= \mathbf{h}^T \mathbf{d}_{R_1}(\mathbf{r}, f), \quad \mathbf{v}_R(\mathbf{r}, f) = \mathbf{h}^T \mathbf{d}_{R_2}(\mathbf{r}, f). \end{aligned}$$

The problem (5) can be reformulated as

$$\begin{aligned} \min_{\mathbf{h} \in \mathbb{R}^{N \times L}} \max_{(\mathbf{r}, f) \in \Omega_M} & \rho_1 |\mathbf{u}_D(\mathbf{r}, f) + j\mathbf{v}_D(\mathbf{r}, f)| \\ & + \rho_2 |\mathbf{u}_R(\mathbf{r}, f) + j\mathbf{v}_R(\mathbf{r}, f)|. \end{aligned}$$

Using the L_1 -norm and introducing the following auxiliary variables:

$$\begin{aligned} \xi_{II,1} &= \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{u}_D(\mathbf{r}, f)|, \quad \xi_{II,2} = \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{v}_D(\mathbf{r}, f)|, \\ \xi_{II,3} &= \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{u}_R(\mathbf{r}, f)|, \quad \xi_{II,4} = \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{v}_R(\mathbf{r}, f)|, \end{aligned}$$

the optimization model corresponding to problem (5) can be established as

Model II

$$\begin{aligned} \min_{\mathbf{h} \in \mathbb{R}^{N \times L}, \boldsymbol{\xi}_{II} \in \mathbb{R}^4} & \boldsymbol{\rho}^T \boldsymbol{\xi}_{II} \\ \text{s.t.} & |\mathbf{u}_D(\mathbf{r}, f)| \leq \xi_{II,1}, \quad \forall (\mathbf{r}, f) \in \Omega_M, \\ & |\mathbf{v}_D(\mathbf{r}, f)| \leq \xi_{II,2}, \quad \forall (\mathbf{r}, f) \in \Omega_M, \\ & |\mathbf{u}_R(\mathbf{r}, f)| \leq \xi_{II,3}, \quad \forall (\mathbf{r}, f) \in \Omega_M, \\ & |\mathbf{v}_R(\mathbf{r}, f)| \leq \xi_{II,4}, \quad \forall (\mathbf{r}, f) \in \Omega_M, \end{aligned} \tag{12}$$

where $\xi_{II} = [\xi_{II,1} \ \xi_{II,2} \ \xi_{II,3} \ \xi_{II,4}]^T$, and $\rho = [\rho_1 \ \rho_1 \ \rho_2 \ \rho_2]^T$ are the weighting parameters developed in (5). In this formulation, the total reverberation are introduced directly into the objective function in order to achieve a better suppression.

For the multi-criteria problem (6), we can divide the reverberation responses $\mathbf{R}_R(\mathbf{r}, f)$ further into early reflections and late reverberation. Rearranging them as

$$\begin{aligned}\mathbf{R}_E^T(\mathbf{r}, f)\mathbf{H}(\mathbf{h}, f) &= \mathbf{h}^T \mathbf{d}_E(\mathbf{r}, f), \\ \mathbf{R}_L^T(\mathbf{r}, f)\mathbf{H}(\mathbf{h}, f) &= \mathbf{h}^T \mathbf{d}_L(\mathbf{r}, f),\end{aligned}$$

expanding the complex functions

$$\begin{aligned}\mathbf{d}_E(\mathbf{r}, f) &= \mathbf{d}_{E_1}(\mathbf{r}, f) + j\mathbf{d}_{E_2}(\mathbf{r}, f), \\ \mathbf{d}_L(\mathbf{r}, f) &= \mathbf{d}_{L_1}(\mathbf{r}, f) + j\mathbf{d}_{L_2}(\mathbf{r}, f),\end{aligned}$$

we have the following additional variables in the objective function:

$$\begin{aligned}\mathbf{u}_E(\mathbf{r}, f) &= \mathbf{h}^T \mathbf{d}_{E_1}(\mathbf{r}, f), \quad \mathbf{v}_E(\mathbf{r}, f) = \mathbf{h}^T \mathbf{d}_{E_2}(\mathbf{r}, f). \\ \mathbf{u}_L(\mathbf{r}, f) &= \mathbf{h}^T \mathbf{d}_{L_1}(\mathbf{r}, f), \quad \mathbf{v}_L(\mathbf{r}, f) = \mathbf{h}^T \mathbf{d}_{L_2}(\mathbf{r}, f).\end{aligned}$$

Introduce the auxiliary variables as

$$\begin{aligned}\xi_{III,1} &= \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{u}_D(\mathbf{r}, f)|, \quad \xi_{III,2} = \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{v}_D(\mathbf{r}, f)|, \\ \xi_{III,3} &= \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{u}_E(\mathbf{r}, f)|, \quad \xi_{III,4} = \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{v}_E(\mathbf{r}, f)|, \\ \xi_{III,5} &= \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{u}_L(\mathbf{r}, f)|, \quad \xi_{III,6} = \max_{(\mathbf{r}, f) \in \Omega_M} |\mathbf{v}_L(\mathbf{r}, f)|,\end{aligned}$$

then the optimization model for problem (6) can be established as

Model III

$$\begin{aligned}\min_{\mathbf{h} \in \mathbb{R}^{N \times L}, \xi_{III} \in \mathbb{R}^6} \quad & \rho^T \xi_{III} \\ \text{s.t.} \quad & \mathbf{U}(\mathbf{r}, f) \leq \xi_{III}, \quad \forall (\mathbf{r}, f) \in \Omega_M,\end{aligned}\tag{13}$$

where

$$\begin{aligned}\mathbf{U}(\mathbf{r}, f) &= [|\mathbf{u}_D(\mathbf{r}, f)| \ |\mathbf{v}_D(\mathbf{r}, f)| \ |\mathbf{u}_E(\mathbf{r}, f)| \\ &\quad |\mathbf{v}_E(\mathbf{r}, f)| \ |\mathbf{u}_L(\mathbf{r}, f)| \ |\mathbf{v}_L(\mathbf{r}, f)|]^T,\end{aligned}$$

and $\xi_{III} = [\xi_{III,1} \ \dots \ \xi_{III,6}]^T$, $\rho = [\rho_1 \ \dots \ \rho_6]^T$. In this formulation, the effect of the early reflections and the late reverberation are separated in the objective function and the corresponding impulse response functions are used in the calculations.

The above multi-criteria models Model II and Model III can also be transformed into linear programming problems similar to (11) of the Model I. Linear programming has been studied extensively in the

literature and the computation for a linear problem is much less than for nonlinear ones [35]. Interior point method [36] can be employed as an efficient solver. According to the formulation, the Model I can be implemented efficiently with just one linear program, while for Model II and Model III, it will take need to solve several linear programs to develop the Pareto optima. Furthermore, Model II will have a relatively lower complexity than Model III because the reverberation effect is considered as a whole in the optimization problem, while it is attempted to differentiate between early and late reverberation in Model III.

IV. EVALUATION OF THE PROPOSED METHOD

In this section, we want to evaluate the performances of the designed beamformers, not only in simulated rooms but also in an actual enclosure. First, one simple rectangular room with adjustable parameters is defined for the fast-ISM room simulator to estimate the RIRs. Then, the proposed models are used to design beamformers and various performance measures are evaluated. A recorded signal is also employed as an input to evaluate the performance in the stopband. Finally, a set of signals are recorded in an actual room to evaluate the beamformers on both dereverberation and noise suppression.

Example 1. Indoor beamformer design

In this example, we define a simple $8m \times 4m \times 3m$ rectangular office room with adjustable absorption coefficients characterizing the room surface, which is also measured by reverberation time T_{60} for convenience. To define a microphone array for beamforming, an equispaced linear array with 5 microphone elements are setup at $\{(2, 3.9, 1.5), (2, 3.95, 1.5), (2, 4, 1.5), (2, 4.05, 1.5), (2, 4.1, 1.5)\}$ in meter, respectively. The array spacing is $0.05m$ to avoid spatial aliasing for the frequency of interest, and a 10-tap FIR filter is used in every channel. We define the passband region as

$$\Omega_p = \{(\mathbf{r}, f) | x = 1m, |y - 4| \leq 0.4m, \\ z = 1.5m, 0.5kHz \leq f \leq 1.5kHz\},$$

and the stopband regions as

$$\Omega_s = \{(\mathbf{r}, f) | x = 1m, |y - 4| \leq 0.4m, \\ z = 1.5m, 2.0kHz \leq f \leq 4.0kHz\}, \\ \cup \{(\mathbf{r}, f) | x = 1m, 1.5m \leq |y - 4| \leq 3.0m, \\ z = 1.5m, 0.5kHz \leq f \leq 1.5kHz\}, \\ \cup \{(\mathbf{r}, f) | x = 1m, 1.5m \leq |y - 4| \leq 3.0m, \\ z = 1.5m, 2.0kHz \leq f \leq 4.0kHz\}.$$

For illustration, the configuration is depicted in Fig. 2.

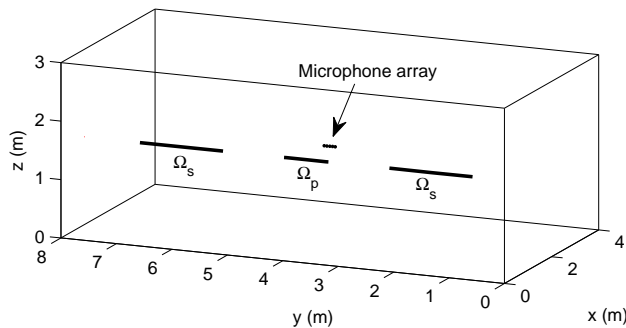


Fig. 2: Setup of acoustic room and layout of microphone array for the indoor beamformer design.

In addition, the desired response function in the passband region Ω_p is given by

$$G_d(\boldsymbol{\lambda}, \boldsymbol{r}, f) = e^{-j2\pi f \frac{\|\boldsymbol{r} - \boldsymbol{r}_c\|}{c}},$$

where c is the sound speed and \boldsymbol{r}_c is the center element of the microphone array which is usually defined as the beamformer output point. In the stopband region Ω_s , we simply set $G_d(\boldsymbol{r}, f) = 0$ in order to filter out the interference and noise. The space-frequency domain of the passband and stopband is discretized by a grid of 30×30 for calculations. A denser grid 120×120 is employed to verify the beamforming performances.

In a typical room environment, when a sound wave strikes a surface, a certain fraction of energy is absorbed and the fractional loss is characterized by the absorption coefficients. The overall effect can be captured by the reverberation time T_{60} (see [25] for more details). Therefore, we use a uniform absorption coefficient for all room boundaries to generate the room acoustics and we choose different T_{60} in the following numerical experiments.

The implementation of Model I is a standard linear programming problem and the beamformer coefficients can be solved directly, but the beamformers designed from the multi-criteria decision Model II and Model III are governed by the weighting parameters $\boldsymbol{\rho}$. In general, for independent objectives, there does not exist a single solution that simultaneously optimizes all objectives in the multi-criteria optimization. The Pareto optima set is therefore needed. For Model II and Model III, we can vary the weighting parameters $\boldsymbol{\rho}$ to obtain the Pareto optima set corresponding to different indoor beamformer designs. In the following, we choose the Pareto optimal solutions based on the total reverberation suppression for Model II and Model III.

To explain the multi-criteria Model III, we give an example to illustrate the conflicting of the early reflections suppression and late reverberation suppression in Fig. 3. Clearly we cannot find a design to

minimize both early reflections and late reverberation and we need to trade off between both suppressions. It demonstrates that the separation scheme for suppressing early reflections and late reverberation in Model III is meaningful.

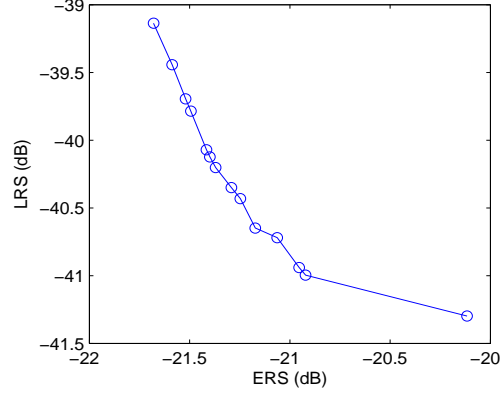


Fig. 3: Conflicting nature of early reflections suppression (ERS) and late reverberation suppression (LRS) with $T_{60} = 0.2s$.

Since all the indoor beamformer design models can achieve the desired response in the passband, the other measures of performance are the noise reduction and reverberation suppression. Therefore, we use the following performance indicators to measure and compare among different models: the direct path response suppression (DRS (in dB)):

$$\frac{1}{\|\Omega_s\|} \sum_{(\mathbf{r}, f) \in \Omega_s} \|\mathbf{R}_D^T(\mathbf{r}, f)\mathbf{H}(\mathbf{h}, f)\|_2,$$

the early reflections suppression (ERS (in dB)):

$$\frac{1}{\|\Omega_s\|} \sum_{(\mathbf{r}, f) \in \Omega_s} \|\mathbf{R}_E^T(\mathbf{r}, f)\mathbf{H}(\mathbf{h}, f)\|_2,$$

and the late reverberation suppression (LRS (in dB)):

$$\frac{1}{\|\Omega_s\|} \sum_{(\mathbf{r}, f) \in \Omega_s} \|\mathbf{R}_L^T(\mathbf{r}, f)\mathbf{H}(\mathbf{h}, f)\|_2.$$

Define the overall performance of the designed beamformer to be the total reverberation suppression (TRS=DRS+ERS+LRS), the numerical results are summarised in Table I under 8 different reverberation time $T'_{60}s = \{0.05s, 0.1s, \dots, 0.4s\}$.

From the results in Table I, the beamformers designed by the proposed models are all effective for noise reduction and reverberation suppression for different reverberation time T_{60} 's. The beamformers designed using the multi-criteria decision based Model II and Model III have better performances than Model I in all performance indicators, and it can be seen that Model II and Model III have more significant

TABLE I: Summary of the performances of the designed indoor beamformers with different reverberation times on desired response, early reflections and late reverberation suppression (in dB).

T_{60}	Model I				Model II				Model III			
	DRS	ERS	LRS	TRS	DRS	ERS	LRS	TRS	DRS	ERS	LRS	TRS
0.05	-14.1400	-39.2890	-61.2107	-14.1267	-21.4810	-45.1816	-65.9210	-21.4624	-22.4063	-44.4467	-67.2945	-22.3791
0.1	-14.1377	-24.4809	-55.4227	-13.7536	-20.5304	-29.4417	-59.1716	-20.0049	-21.6910	-29.8507	-59.7177	-21.0730
0.15	-14.1377	-19.4207	-47.1581	-13.0090	-20.4440	-21.9449	-41.0765	-18.0977	-21.6191	-24.7314	-50.3162	-19.8880
0.2	-14.1364	-16.6324	-38.2722	-12.1865	-20.2677	-21.1781	-38.8466	-17.6556	-20.8478	-21.2104	-40.7866	-17.9921
0.25	-14.1353	-14.7239	-32.0293	-11.3718	-18.8318	-17.6315	-32.0034	-15.0907	-19.0192	-18.1811	-34.5490	-15.5151
0.3	-14.1345	-13.3163	-27.4316	-10.6047	-17.0809	-15.2531	-27.7305	-12.9155	-18.5007	-16.4701	-31.0600	-14.2657
0.35	-14.1337	-12.2411	-23.9777	-9.9015	-17.7867	-14.5372	-23.8327	-12.5210	-18.6589	-15.4044	-27.1069	-13.5285
0.4	-14.1331	-11.3921	-21.2057	-9.2533	-17.4041	-13.4633	-21.6259	-11.5425	-18.3217	-13.2024	-23.5995	-11.7445

TABLE II: Summary of off-design performances (in dB) of the indoor beamformers designed for a room with a reverberation time of $T_{60} = 0.1s$.

T_{60}	Model I				Model II				Model III			
	DRS	ERS	LRS	TRS	DRS	ERS	LRS	TRS	DRS	ERS	LRS	TRS
0.08	-14.1395	-28.1768	-57.3646	-13.9712	-21.1037	-33.8798	-62.1626	-20.8800	-20.9454	-33.4870	-62.0590	-20.7097
0.09	-14.1392	-26.1214	-56.0164	-13.8722	-22.0220	-31.7308	-60.2814	-21.5802	-21.8461	-31.6153	-60.6142	-21.4102
0.11	-14.1387	-23.1337	-54.5959	-13.6228	-20.7166	-28.2511	-57.8366	-20.0103	-22.0703	-28.5929	-58.3538	-21.1962
0.12	-14.1385	-22.0046	-53.4133	-13.4806	-20.5373	-26.9751	-56.1224	-19.6475	-21.2157	-26.9252	-56.7525	-20.1816

improvements on DRS than Model I. Also, beamformers designed from Model III performs slightly better than that from Model II in the TRS, which gives credit to the flexibility introduced into Model III in choosing different weights. To take a closer look at the the overall performance of the designed beamformers from Model III, frequency response functions with reverberation time $T_{60} = 0.1s$ and $T_{60} = 0.2s$ are shown in the Fig. 4 (in these figures, $x(m)$ and $f(Hz)$ denotes the space and the frequency domain, respectively).

In addition, we use the perturbed room acoustics to examine the robustness of the designed beamformers. We apply the designed beamformers at reverberation time $T_{60} = 0.1s$ into the perturbed acoustic room with reverberation time $T_{60} = \{0.08s, 0.09s, 0.11s, 0.12s\}$, respectively. From the experimental results summarized in Table II, the designed beamformers from the proposed models are all robust to small perturbations in the reverberation time.

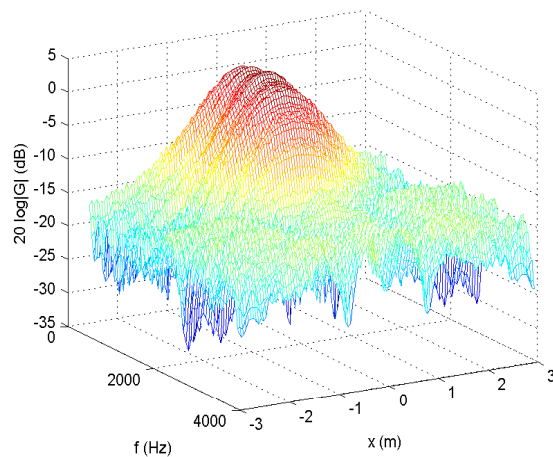
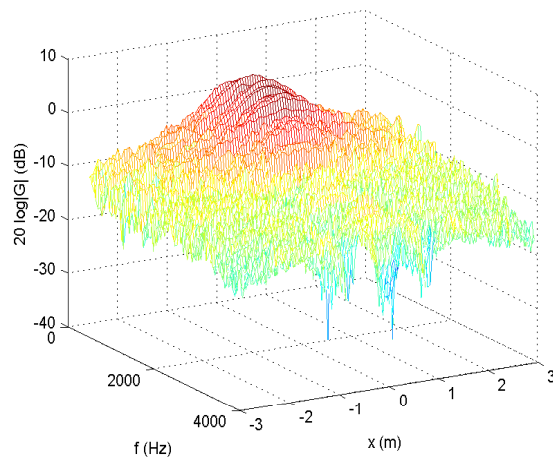
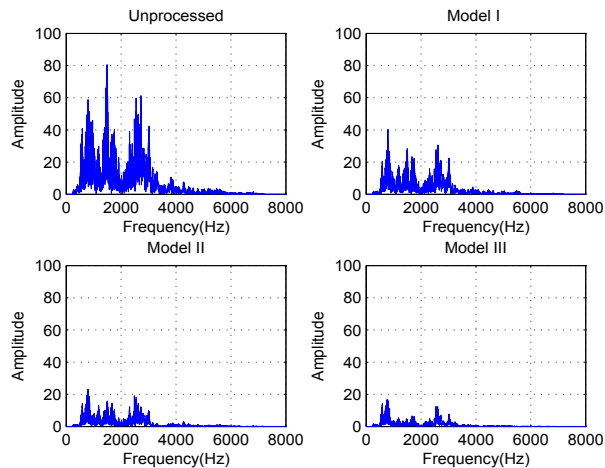
(a) $T_{60} = 0.1s$ (b) $T_{60} = 0.2s$

Fig. 4: Overall performance of the indoor beamformers designed by using Model III.

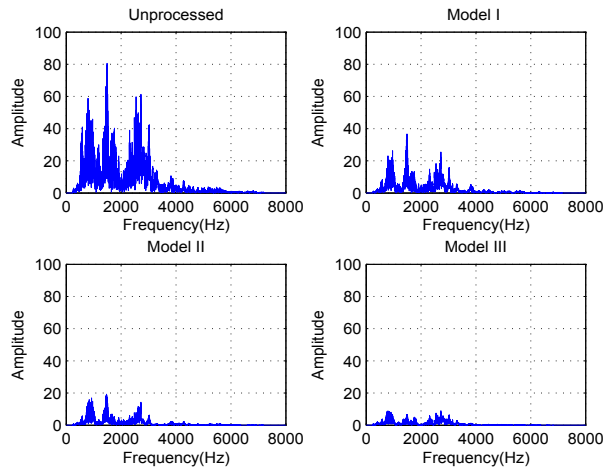
Example 2. Signal suppression in stopband

In this example, we input a test signal in the stopband and evaluate the performances of the designed indoor beamformers. We use a male speech as the input signal and employ the setup of the room and layout of microphone array defined in Example 1 with reverberation time $T_{60} = 0.1s$ for the simulation. The beamformers are designed using Model I, Model II and Model III. For comparison, we use the signal captured by the center element of the microphone array as a reference signal. The positions tested are at $\mathbf{r} = (1, 2, 1.5) \in \Omega_s$ and $\mathbf{r} = (1, 1, 1.5) \in \Omega_s$ and the amplitude suppressions are depicted in Fig. 5.

It can be seen that all the designed beamformers are effective on the noise suppression in the stopband, and the beamformer designed using Model III is slightly better than Model I and Model II.



(a) Noise at $\mathbf{r} = (1, 2, 1.5)$



(b) Noise at $\mathbf{r} = (1, 1, 1.5)$

Fig. 5: Performances of the designed indoor beamformers in the stopband.

There are many objective measurement methods derived in the literature to evaluate the performance of speech enhancement [37], such as the signal-to-noise ratio (SNR) [38] and signal-to-interference ratio (SIR) [39]. In order to measure the noise and reverberation suppression in the stopband, we can similarly define the overall suppression measurement as follows:

$$Supp = 10 \log_{10} \frac{\|\mathbf{S}(f)\|_2^2}{\|\mathbf{Y}(f)\|_2^2} \quad (\text{dB}), \quad (14)$$

where $S(f)$ and $Y(f)$ represent the frequency spectrums of the input signal and the beamformer output, respectively. We select the stopband location from $\mathbf{r} = (1, 1, 1.5)$ to $\mathbf{r} = (1, 2.5, 1.5)$ with $0.1m$ spacing to evaluate the noise suppression. A summary of the results is depicted in Fig. 6. We can see that all the suppression measurements in the stopband interval $[2.5m, 1m]$ have been enhanced after beamforming by using the proposed models, and the farther away from passband, the better performance on suppression. Moreover, we can see that the multi-criteria based Model II and Model III perform better than Model I; in particular, the proposed Model III performs best among all the methods.

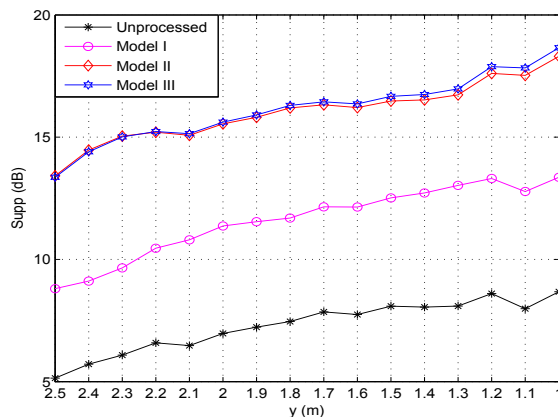


Fig. 6: Evaluation of designed indoor beamformers in the stopband.

Example 3. Performance evaluation with recordings

In this example, the effectiveness of the proposed models for beamformer design is evaluated using sound recordings in an actual meeting room. We also compare the performances of beamformers designed via the matched-filter array (MFA), direct path transfer function and the proposed methods. We choose a meeting room to generate the room acoustics with the parameter settings of the environment and microphone array as shown in Table III.

TABLE III: Parameter settings for the experiment.

Room dimensions	$8.015m \times 4.671m \times 2.7m$
Temperature	22.6°C
Relative humidity	39%
Reverberation time	$T_{60} \approx 0.45s$
Plan view	Fig. 7

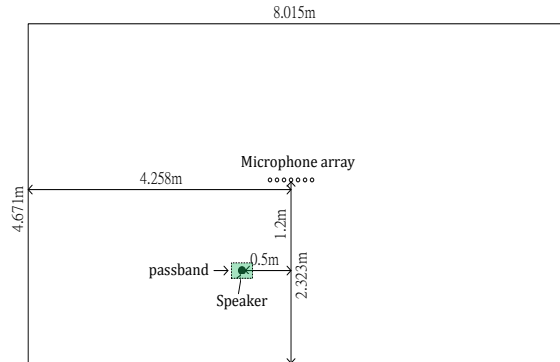


Fig. 7: Setup of the meeting room and layout of the microphone array.

In order to setup a microphone array, a linear microphone array composed of 7 elements having an inter-element spacing of $0.04m$ is placed in the middle of the table, with about $1.13m$ above the floor. Then we set the passband to be the rectangular region $\Omega_p = \{x_0 \pm 0.05m, y_0 \pm 0.05m\}$ with a 5×3 grid for discretization, where (x_0, y_0) is the speaker position and they are all $1.13m$ above the floor. For the recordings, a male and a female speech are played back at $48kHz$ for about 9 seconds. The recording signals are re-sampled to be $16kHz$ for beamforming processing. In the frequency domain, we set the passband and stopband to be $[0.5kHz, 4kHz]$ and $[5kHz, 16kHz]$, respectively, and with 40 grid points for discretization.

We define a 20-tap FIR filter behind each microphone element to design the beamformer, and select microphone element at the centre as the beamformer output. In general, the direct path captured signals are defined as the desired output and used in performance evaluation. As the recordings are contaminated by reflections and reverberation, therefore, we simulate the desired output signals by using the convolution of original speeches with the direct path impulse responses and extract the components in frequency region $[0.5kHz, 4kHz]$ to generate the desired signals.

To evaluate the performance of dereverberation, we introduce the segmental signal-to-reverberation ratio (SRR) [40] as

$$SRR_{seg} = \frac{1}{Q} \sum_{q=0}^{Q-1} 10 \log_{10} \frac{\|\mathbf{S}(q)\|^2}{\|\mathbf{S}(q) - \hat{\mathbf{S}}(q)\|^2} \text{ dB}, \quad (15)$$

where $\mathbf{S}(q)$ and $\hat{\mathbf{S}}(q)$ represent the desired speech signal estimated according to above method and the enhanced speech from the beamformer out in q -th time frame, respectively.

In the experiment, we first measure the room acoustics to characterize the meeting room, then we design a similar room using the room simulator with similar characteristics. We take the reverberation time as $T_{60} = 0.45s$ for the fast-ISM room simulator to estimated the corresponding RIRs during the

design of beamformer. Then, we use the proposed implementation models to design beamformers and carry out beamforming for speech quality enhancement. In addition, we apply the matched-filtering array (MFA) technique to process the reverberant signals for comparison, and also apply the beamformer designed from the direct path transfer function. The experimental results are summarized in Table IV, in which the column ‘Reverberant’ is the unfiltered reverberant signal captured by the centering element of microphone array, ‘MFA-method’ is the filtered signal using the matched-filter array, and the ‘Direct path’, ‘Model I’, ‘Model II’ and ‘Model III’ are the filtered signals using beamformers designed via the direct path transfer function, Model I, Model II and Model III, respectively.

TABLE IV: Comparison results among proposed models on the overall dereverberation performance SRR_{seg} .

Speech	Reverberant	MFA-method	Direct path	Model I	Model II	Model III
Male	-6.7793	-6.4732	-5.6748	-4.9231	-4.2572	-3.9258
Female	-7.3200	-6.4666	-5.8230	-5.1694	-4.5500	-4.1520

TABLE V: Comparison results among proposed models on the stopband noise suppression $Supp$.

Speech	MFA-method	Direct path	Model I	Model II	Model III
Male	0.8868	3.1087	5.5757	7.2275	9.2821
Female	1.2224	4.8809	6.5341	8.1541	9.2587

According to the evaluation in Table IV, we can see that the qualities of Male and Female speeches have been deteriorated by the room acoustics. After beamforming, all the segmental SRR scores are increased by different levels, and the results show that all the beamformers designed are effective. In addition, we can see that the beamformer based on Model III has the best performance on dereverberation.

Furthermore, we also evaluate the stopband suppression for these methods by using the indicator $Supp$ defined in (14) and the results are in Table V. From the stopband suppression measurement, the beamformers designed based on the MFA-method and direct path transfer function have very poor performance on both Male or Female speeches, whereas all the other proposed models perform better. Among all, Model III shows the best result.

To illustrate further the overall beamforming performance, we plot the Female speeches before and after processing by using the beamformers in Fig. 8. We also plot the spectrograms by using a 2 second signals in Fig. 9. From the results, reflections of the speeches are removed to a certain extent in the outputs

of the designed beamformers from the MFA-method, direct path method, and the proposed methods. It is further observed that the proposed models are more effective for dereverberation. Moreover, it can be seen from the spectrograms that distortion is not serious.

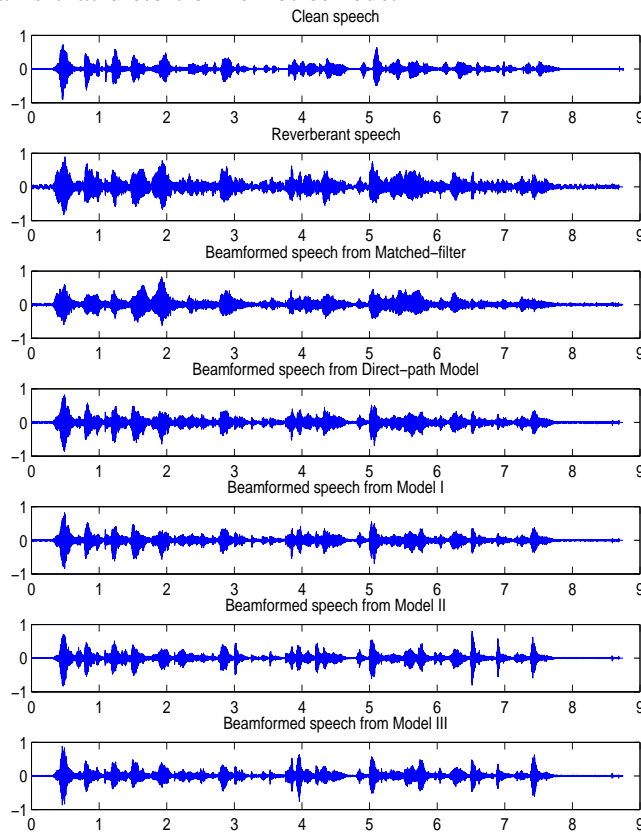
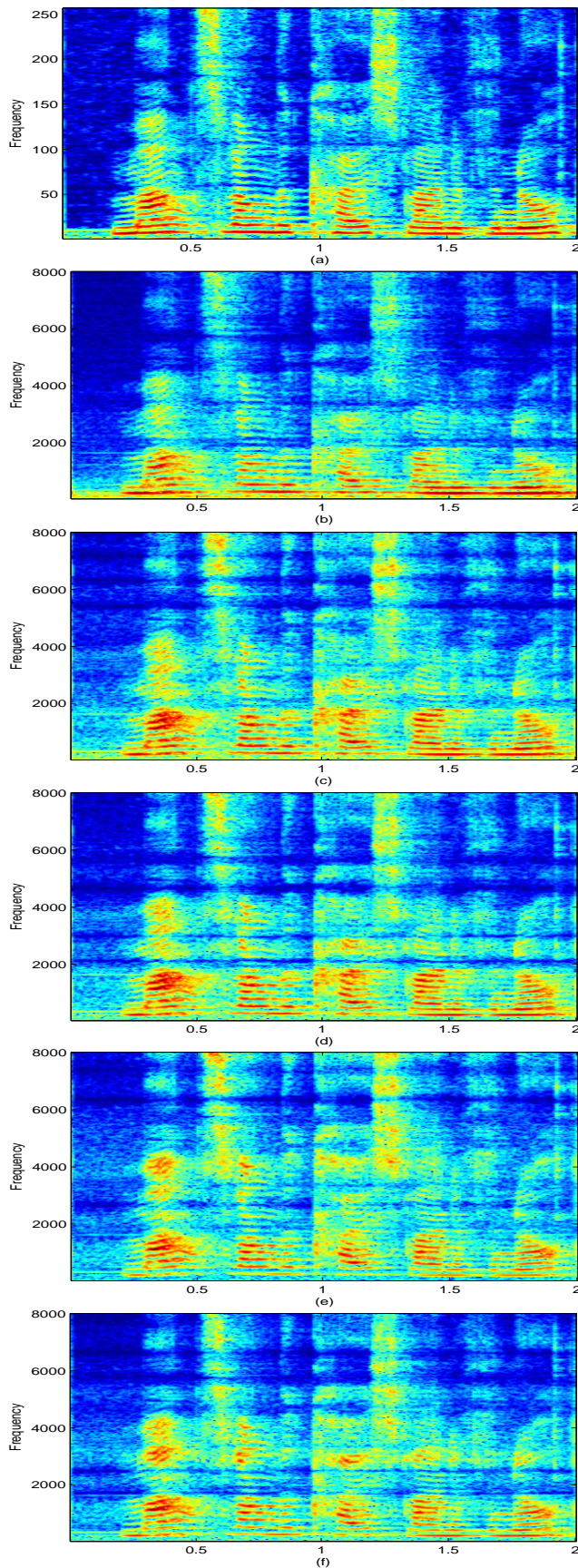


Fig. 8: Beamforming performance on the Female speech.

V. CONCLUSION

In this paper, we have formulated the indoor beamformer design problem into a minimax semi-infinite programming problem. We introduced the fast-ISM room simulator to estimate the RIRs, and separated them into a direct path response, early reflections and late reverberation. Based on our simulated results, we have demonstrated that early and late reverberation behave in a Pareto manner. We went on to propose and construct two multi-criteria optimization models incorporating separated RIRs for the design of more effective beamformers in a reverberant environment.

To verify the effectiveness of the proposed indoor beamformer designs, we have employed both simulated scenarios and actual recordings in a meeting room. The experimental results showed that multi-criteria based models can improve the overall performance. The designs are also robust to small perturbations to the room acoustics. Furthermore, we have evaluated the performance using actual record-



May 26, 2014

DRAFT

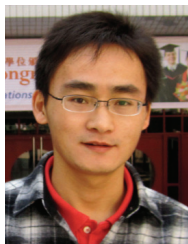
Fig. 9: Female speech spectrograms: (a) Clean speech; (b)-(f) Beamformed speeches from 'MFA-method', 'Direct path', 'Model I', 'Model II' and 'Model III'.

ings in a meeting room, and have shown that the proposed models all effective in dereverberation and noise suppression. In particular, the proposed Model III has the best performance. As a future work, it would also be of interest to investigate the placement problem [41] in dereverberation, and the use of calibration signals to estimate RIRs and employed in the multi-criteria optimization.

REFERENCES

- [1] T. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, Englewood Cliffs, NJ, 1996.
- [2] J. Liberti and T. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*. Prentice Hall, Upper Saddle River, NJ, 1999.
- [3] F. Gross, *Smart Antennas for Wireless Communications: With MATLAB*. McGraw-Hill, New York, 2005.
- [4] P. Naylor and N. Gaubitch, *Speech Dereverberation*. Springer-Verlag, London, 2010.
- [5] A. Levy, S. Gannot, and E. Habets, "Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1540–1555, 2011.
- [6] D. Ward and G. Elko, "Mixed nearfield/farfield beamforming: a new technique for speech acquisition in a reverberant environment," in *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1997.
- [7] J. Flanagan, A. Surendran, and E. Jan, "Spatially selective sound capture for speech and audio processing," *Speech Communication*, vol. 13, no. 1-2, pp. 207–222, 1993.
- [8] E. Jan, P. Svaizer, and J. Flanagan, "Matched-filter processing of microphone array for spatial volume selectivity," in *IEEE International Symposium on Circuits and Systems*, vol. 2, 1995, pp. 1460–1463.
- [9] F. MacWilliams and N. Sloane, "Pseudo-random sequences and arrays," *Proceedings of the IEEE*, vol. 64, no. 12, pp. 1715–1729, 1976.
- [10] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1614–1626, 2001.
- [11] S. Gannot and I. Cohen, "Speech enhancement based on the general transfer function GSC and postfiltering," *Speech and Audio Processing, IEEE Transactions on*, vol. 12, no. 6, pp. 561–571, 2004.
- [12] R. Talmon, I. Cohen, and Gannot, "Relative transfer function identification using convolutive transfer function approximation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 4, pp. 546–555, 2009.
- [13] Y. Huang, J. Benesty, and J. Chen, "Adaptive blind multichannel identification," in *Springer Handbook of Speech Processing*, J. Benesty, M. Sondhi, and Y. Huang, Eds. New York: Springer-Verlag, 2007, ch. 13, pp. 259–282.
- [14] D. Botteldoore, "Finite-difference time-domain simulation of low-frequency room acoustic problems," *Journal of the Acoustical Society of America*, vol. 98, no. 6, pp. 3302–3308, 1995.
- [15] L. Wrobel, *The boundary element method, Applications in Thermo-Fluids and Acoustics*. Wiley, 1991.
- [16] M. Vorländer, *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*. Springer-Verlag, Berlin, Heidelberg, 2008.
- [17] H. Lehnert and J. Blauert, "Principles of binaural room simulation," *Appl. Acoust.*, vol. 36, no. 3-4, pp. 259–291, 1992.
- [18] U. Svensson, "Modelling acoustic spaces for audio virtual reality," in *Workshop on Model based Processing and Coding of Audio (MPCA)*, Leuven, Belgium, 2002, pp. 109–116.

- [19] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.
- [20] J. Borish, "An extension of the image model to arbitrary polyhedra," *J. Acoust. Soc. Amer.*, vol. 75, pp. 1827–1836, 1984.
- [21] M. Schroeder and K. Kuttruff, "On frequency response curves in rooms: Comparison of experimental, theoretical, and monte carlo results for the average frequency spacing between maxima," *J. Acoust. Soc. Amer.*, vol. 34, no. 1, pp. 76–80, 1962.
- [22] H. Kuttruff and E. Mommertz, "Room acoustics," in *Handbook of Engineering Acoustics*, G. Müller and M. Möser, Eds. Springer Berlin Heidelberg, 2013, pp. 239–267.
- [23] C. Christensen, G. Koutsouris, and J. Rindel, "The ISO 3382 parameters: Can we simulate them? Can we measure them?" in *International Symposium on Room Acoustics*, 2013.
- [24] I. Titze, *Principles of voice production*. Prentice Hall Englewood Cliffs, 1994.
- [25] E. Lehmann and A. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *Trans. Audio, Speech and Lang. Proc.*, vol. 18, pp. 1429–1439, 2010.
- [26] H. Kuttruff, *Room Acoustics, 4th ed.* London: Spon Press, 2000.
- [27] E. Habets, I. Cohen, S. Gannot, and P. Sommen, "Joint dereverberation and residual echo suppression of speech signals in noisy environments," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1433–1451, 2008.
- [28] K. Yiu, N. Grbic, T. Kok-Lay, and S. Nordholm, "A new design method for broadband microphone arrays for speech input in automobiles," *IEEE Signal Processing Letters*, vol. 9, no. 7, pp. 222–224, 2002.
- [29] K. Yiu, X. Yang, S. Nordholm, and K. Teo, "Near-field broadband beamformer design via multidimensional semi-infinite linear programming techniques," *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 6, pp. 725–732, 2003.
- [30] Z. Feng, K. Yiu, and S. Nordholm, "A two-stage method for the design of near-field broadband beamformer," *IEEE Trans. Signal Processing*, vol. 59, no. 8, pp. 3647–3656, 2011.
- [31] R. Fletcher, *Practical Methods of Optimization*. New York: Wiley, 1987.
- [32] R. Hettich and K. Kortanek, "Semi-infinite programming: Theory, methods, and applications," *SIAM Review*, vol. 35, no. 3, pp. 380–429, 1993.
- [33] R. Reemsten and S. Górnér, *Numerical methods for semi-infinite programming: A survey*. Kluwer Academic, Boston, 1998.
- [34] K. Teo, X. Yang, and L. Jennings, "Computational discretization algorithms for functional inequality constrained optimization," *Annals of Operations Research*, vol. 98, no. 1-4, pp. 215–234, 2000.
- [35] G. H. Golub and C. F. van Loan, *Matrix Computations, 4th edition*. Johns Hopkins University Press, 2012.
- [36] D. Luenberger and Y. Ye, *Linear and Nonlinear Programming, 3rd edition*. Springer, 2008.
- [37] P. Loizou, *Speech Enhancement: Theory and Practice*. CRC Press, Boca Raton, FL, 2007.
- [38] J. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in *Proc. Internat. Conf. on Spoken Language Processing*, San Francisco, CA, 1998, pp. 2819–2822.
- [39] K. Hild, D. Erdogmus, and J. Principe, "Experimental upper bound for the performance of convolutive source separation methods," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 627–635, 2006.
- [40] P. Naylor, E. Habets, J.-C. Wen, and N. Gaubitch, "Models, measurement and evaluation," in *Speech Dereverberation*, P. Naylor and N. Gaubitch, Eds. Springer, 2010, pp. 21–56.
- [41] Z. Feng, K. Yiu, and S. Nordholm, "Placement design of microphone arrays in near-field broadband beamformers," *IEEE Trans. Signal Processing*, vol. 60, no. 3, pp. 1195–1204, 2012.



Zhibao Li received the B.S. degree in mathematics and applied mathematics from the Anhui Normal University, Wuhu, China, in 2007, the M.S. degree in Computational Mathematics from the Nanjing Normal University, Nanjing, China, in 2010, and the Ph.D degree in applied mathematics from The Hong Kong Polytechnic University, Hong Kong, China, in 2014.

He is currently a postdoctoral researcher in the State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. His current research interests include near-field beamformer design, acoustic signal processing, speech enhancement, deep learning, and numerical optimization.



Ka Fai Cedric Yiu received his M.Sc. from University of Dundee and University of London, and D.Phil. from University of Oxford. Over the century, he had worked closely with the industry on different projects in University of Oxford and University College of London. He started his lecturing career in the University of Hong Kong. He is currently working in the Hong Kong Polytechnic University. He serves on the program committee of the conference ASAP for several years and has served on the organizing committee of a number of conferences including FPT02, LSCM06, FERM08. He has also organized a number of special sessions in conferences including ICOTA7, CISP08, ISCE2009, EURO2009. He has published over 90 journal publications and given over 30 conference presentations. He holds a U.S. patent in signal processing. His current research interests include optimization and optimal control, signal processing and sensor array processing.



Sven Nordholm (SM) received his Ph.D. in Signal Processing in 1992, Licentiate of engineering 1989 and MscEE (Civilingenjör) 1983 all from Lund University, Sweden. He is professor at Curtin University, Perth, Australia.

He was one of the founding members of the Department of Signal Processing, Blekinge Institute of Technology, BTH, Sweden in 1990. At BTH he held positions as Lecturer, Senior Lecturer, Associate Professor and Professor. Since 1999 he has been at Curtin University of Technology in Perth, Western Australia. From 1999-2002 he was director of ATRI and Professor at Curtin University. From 2002-2009 he was director Signal Processing Laboratory, WATRI, Western Australian Telecommunication Research Institute, a joint institute between The University of Western Australia and Curtin University. Currently, he is Prof. Signal Processing and Head of Department, Dept. of Electrical and Computer Engineering, Curtin University. He was Chief Scientist and co-founder of a start-up company Sensear providing voice communication in extreme noise conditions. He was Eureka finalist 2013 for that work.

He is the author and co-author of more than 200 peer reviewed papers in conferences and journals. He is a patent holder of 9 patents. He was a Technical Chair of APCC 2005 and is the technical chair for ICCAIS 2013. He was a keynote speaker in ISPACS 2011. He is an associate editor Eurasip Journal on Advances in Signal Processing and Journal of Franklin Institute. He has been a guest editor for special issues on Microphone Arrays and Hearing Aids in Eurasip Journal on Advances in Signal Processing. He is a Senior Member of IEEE and a member of IEEE TC AASP.

His research interests include Assistive Listening Devices, Speech Enhancement, Blind Signal Separation, Adaptive and Optimum Microphone Arrays, Acoustic Echo Cancellation, Adaptive Signal Processing, Sub-band Adaptive Filtering and Filter Design.