Engineering Applications of Artificial Intelligence 47 (2016) 81–87

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Beamformer configuration design in reverberant environments

Zhibao Li, Ka-Fai Cedric Yiu *

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, PR China

ARTICLE INFO

Available online 10 June 2015

Keywords:
Microphone array configuration
Reverberation
LCMV beamforming
Hybrid descent method

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.

1. Introduction

Beamforming techniques have been employed extensively in speech communication systems, teleconferencing, speech recognition, and hearing aids (Benesty et al., 2008; Gannot and Cohen, 2008). 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 Capon (1969) 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 (Frost, 1972; Er and Cantoni, 1983) 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 Feng et al. (2015), unless 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 (Feng et al., 2012). 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 (Mayhan, 1980; Sarajedini, 1999; Oliveri et al., 2009; Rocca and Haupt, 2010); different algorithms have been developed, including evolutionary programming (Kumar et al., 1998; Delgado et al., 2010), genetic algorithm (Haupt, 1994; Chen et al., 2007), simulated annealing algorithm (Trucco and Murino, 1999; Trucco, 2002; Doblinger, 2008) and pattern search algorithm (Razavi and Forooraghi, 2008). For applications inside a vehicle, microphones are restricted to be in several dedicated areas and an evolutionary algorithm has been proposed in Ayllón et al. (2014). In formulating the general multi-dimensional design problem, a nonlinear optimization problem using the L_2-norm was proposed in Feng et al. (2012), 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 (Feng et al., 2012). Also, a hybrid descent method using the genetic algorithm

* Corresponding author.
E-mail addresses: zbli0307@163.com (Z. Li), macyiu@polyu.edu.hk (K.-F.C. Yiu).

http://dx.doi.org/10.1016/j.engappai.2015.04.015
0952-1976/

was developed in Li et al. (2013) 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) (Lehnert and Blauert, 1992; Svensson, 2002; Allen and Berkley, 1979; Borish, 1984), and have been applied for indoor beamformer design (Li and Yiu, 2013; Li et al., 2014). During the configuration design process, the room acoustics must be re-calculated whenever the configuration changes. A fast implementation of the ISM proposed in Lehmann and Johansson (2010) 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 2, we define the configuration design problem by using the estimated RIRs based on the fast image-source method. In Section 3, we study the performance limit on the filter length of the LCMV beamformer by solving the optimal frequency response function directly. In Section 4, after eliminating the filter length effect, we develop a hybrid descent method using the genetic algorithm to tackle the configuration design problem. In Section 5, several numerical examples are given to evaluate the effectiveness of the proposed method. Conclusion is given in Section 6.

2. Microphone array configuration design

Assume that the indoor environment is a rectangular room with N sources settled at $\boldsymbol{r}_n$, $n=0,\dots,N-1$. 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.

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 $\boldsymbol{\lambda}=(\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_M)\in\boldsymbol{\Lambda}\subset\mathbb{R}^{3\times M}$, suppose the frequency domain transfer functions describing the acoustics are denoted by

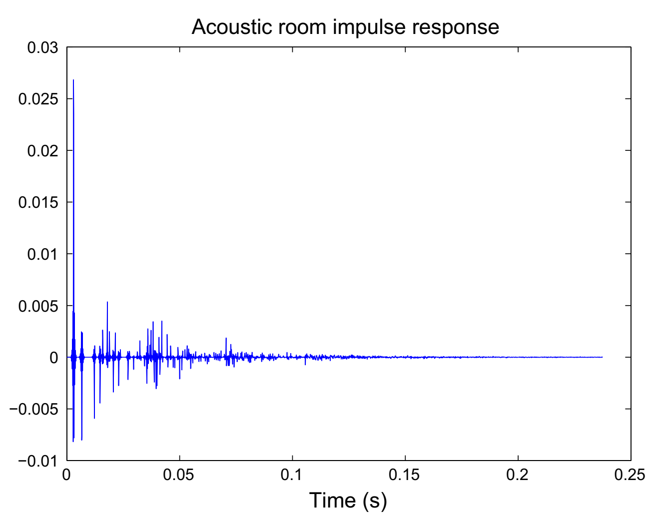

Fig. 1. A typical example of the RIR.

$H_{n,m}(\boldsymbol{\lambda},\omega)$, signals captured by the microphone array are represented by

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

Assume that there is an L-tap FIR filter behind each microphone with coefficients $w_m=[w_m(0),w_m(1),\dots,w_m(L-1)]^T$, $m=1,\dots,M$, 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

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

where $d_0(\omega)$ is defined as

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

and $0\le\tau_L\le L-1$ 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 $\boldsymbol{W}(\omega)=[W_1(\omega)\ \dots\ W_M(\omega)]^T$ denotes the beamformer response vector and $\boldsymbol{X}(\omega)=[X_1(\omega)\ \dots\ X_M(\omega)]^T$ is the received signal.

Notice that the beamformer output $Y(\omega)$ is a function of the microphone array placement $\boldsymbol{\lambda}$ and filter coefficients $\boldsymbol{w}=[w_1, w_2,\dots,w_M]^T$. 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:

$$\min_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda},\boldsymbol{w}\in\mathbb{R}^{M\times L}}\quad \boldsymbol{F}(\boldsymbol{\lambda},\boldsymbol{w})$$
$$\text{s.t.}\quad \|\mathbf{r}_i-\mathbf{r}_j\|^2\ge\varepsilon_0,\quad i,j=1,2,\dots,M,\ i\ne j, \tag{6}$$

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.

3. 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 $W_{m,opt}(\omega)$, $m=1,\dots,M$ are associated with the infinite length filters (Feng et al., 2011) in the following lemma.

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

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

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*