Digital Signal Processing 32 (2014) 100–108

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Design of minimax robust broadband beamformers with optimized microphone positions

R.C. Nongpiur *

Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, V8W 3P6, Canada

ARTICLE INFO

Article history:
Available online 21 May 2014

Keywords:
Optimized sensor positions
Acoustic beamforming
Broadband beamformer
Constrained optimization

ABSTRACT

A new method for the design of robust minimax far-field broadband beamformers with optimized microphone positions is proposed. The method is formulated as an iterative optimization problem where the maximum passband[1] magnitude response error is minimized and the microphone positions are optimized while ensuring that the minimum stopband attenuation is above a prescribed level. To maintain robustness, we constrain a sensitivity parameter, namely, the white noise gain, to be above prescribed levels across the frequency band. An additional feature of the method, which is quite useful in certain applications, is that it provides the capability of constraining the gain in the transition band to always lie below the maximum gain in the passband. Performance comparisons with existing methods show that the optimization of the microphone positions results in beamformers with superior performance.

1. Introduction

Microphone arrays are widely used in speech communication applications such as hands-free telephony, hearing aids, speech recognition, and teleconferencing systems. A technique that is widely used with microphone arrays to enhance a speech signal from a preferred spatial direction is beamforming [1]. In general, the beamforming approach can be fixed or adaptive, depending upon whether the spatial directivity pattern is fixed or varies adaptively on the basis of incoming data. Though adaptive beamforming performs better when the acoustic environment is time-varying, fixed beamforming is preferred in applications where the direction of the sound source is fixed, such as in in-car communication systems [2] or in hearing aids. In addition, fixed beamformers also have lower computational complexity and are easier to implement.

In many beamformer applications, such as in-car communication systems, voice recognition systems, video conferencing systems, etc., there is often a need to ensure that the gain across the passband has little variation from unity while that in the stopband is below a prescribed level. Consequently, for the design of such beamformers a straightforward approach is to formulate the problem as a minimax optimization problem [3]. In applications where high quality speech or audio is desired, a passband with good linear-phase characteristics is usually preferred to minimize signal distortion.

In broadband beamformer design, better frequency invariance can be achieved by arranging the microphone elements non-uniformly in an optimal manner. This is because the more widely separated sensors facilitate better performance at lower frequencies, while the closely spaced ones prevent spatial aliasing at higher frequencies. In [4–8], the microphone elements are arranged in the form of a nested array by appropriately combining several uniformly-spaced sub-arrays. Alternatively, in [9–11], the microphone positions are obtained by approximating a continuously distributed sensor as a discrete set of filtered broadband omnidirectional array elements. In both the methods, the signal from each microphone is appropriately filtered to adjust the time delay and to prevent spatial aliasing, which happens when the wavelength of the signal exceeds twice the distance between the adjacent microphones. However, the microphone positions computed by using the above methods have two serious drawbacks. The first is that the positions are computed on the assumption that the array filters can have any length thereby making their performance sub-optimal when the prescribed filter length is not sufficiently long. The second drawback is the assumption that the array can have any length; such an assumption may not be applicable in certain applications such as in hearing aids and in-car communication systems where there are physical constraints on the array aperture size. Furthermore, as evident from earlier designs for superdirective narrowband arrays [12–16], broadband beamformers

* Fax: +1 250 721 6052.
E-mail address: rnongpiu@ece.uvic.ca.

[1] In this paper, unless explicitly stated, the terms *passband*, *stopband*, and *transition band* refer to the *angular passband*, *angular stopband*, and *angular transition band* of the beamformer, respectively.

http://dx.doi.org/10.1016/j.dsp.2014.05.006
1051-2004/

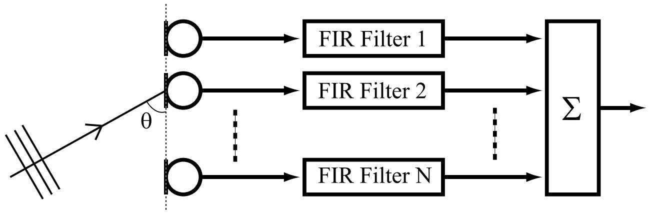

Fig. 1. Filter and sum broadband beamformer.

designed for physically-compact applications can likewise become very sensitive to errors in array imperfections and therefore robustness constraints need to be incorporated in the design [3, 17–23]. In [17–21], the statistics of microphone characteristics are taken into account to derive broadband beamformers that are robust to microphone mismatches, while in [3,22,23] the white noise gain (WNG) is incorporated in the design to ensure that the beamformer is robust to spatial white noise and array imperfections. The use of the WNG constraint is not new and has been used in earlier beamformer designs to ensure robustness in superdirective beamformers [12–14]. More recently, a least-squares approach [24] that uses mixed stochastic and analytic optimization to synthesize both the sensor arrays and filter coefficients for robust wideband beamformers has been proposed. In the approach, a trade-off parameter is provided so that the user can tune the mainlobe width and sidelobe energy of the beamformer.

In this paper, we propose a design method where the maximum passband magnitude response error is minimized and the microphone positions are optimized while ensuring that the minimum stopband attenuation is above a prescribed level. Although the transition region is usually treated as a “don't care” region, in many practical applications excessive gain in this region could be undesirable. Our method provides the capability of controlling the gain in transition bands so that it does not exceed the maximum gain in the passband. To maintain robustness, we constrain a sensitivity parameter, namely, the white noise gain, to be above prescribed levels across the frequency band. The method is formulated as an iterative second-order cone programming problem (SOCP) as was done for the design of nearly linear-phase beamformers [3] and IIR filters [25,26]. Numerical results on various types of beamformers show that the proposed method, with optimized microphone positions, results in beamformers with much lower maximum passband ripple for the same stopband attenuation when compared with beamformers where the microphone positions are fixed and not optimized.

The paper is organized as follows. In Section 2, we describe the filter-and-sum beamformer and the associated error formulations of the beamformer response and WNG for a linear array in far-field. Then in Section 3, we develop formulations for solving the optimization problem. In Section 4, performance comparisons between the proposed method and two existing methods for computing the microphone positions are carried out. Conclusions are drawn in Section 5.

2. Far-field broadband beamforming

In this paper, we assume a far-field signal impinging on a linear microphone array that is realized as a filter-and-sum beamformer, as shown in Fig. 1. The microphones are assumed to be omnidirectional and the filters are FIR. If N is the number of microphones and L is the length of each filter, the response of the filter-and-sum beamformer is given by [3]

$$B(\mathbf{x},\mathbf{d},\omega,\theta)=\sum_{n=0}^{N-1}\hat{\mathbf{g}}(d_n,\omega,\theta)^T\mathbf{x}_n=\mathbf{g}(\mathbf{d},\omega,\theta)^T\mathbf{x} \tag{1}$$

where

$$\mathbf{x}^T=\left[\mathbf{x}_0^T\,\mathbf{x}_1^T\,\cdots\,\mathbf{x}_{N-1}^T\right] \tag{2}$$

$$\mathbf{d}^T=[d_0\,d_1\,\cdots\,d_{N-1}] \tag{3}$$

$$\mathbf{g}(\mathbf{d},\omega,\theta)^T=\left[\hat{\mathbf{g}}(d_0,\omega,\theta)^T\,\hat{\mathbf{g}}(d_1,\omega,\theta)^T\,\cdots\,\hat{\mathbf{g}}(d_{N-1},\omega,\theta)^T\right] \tag{4}$$

$$\mathbf{x}_n=[x_{n,0}\,x_{n,1}\,\cdots\,x_{n,L-1}]^T \tag{5}$$

$$\hat{\mathbf{g}}(d,\omega,\theta)=\left[g_0(d,\omega,\theta)\,g_1(d,\omega,\theta)\,\cdots\,g_{L-1}(d,\omega,\theta)\right]^T \tag{6}$$

$$g_l(d,\omega,\theta)=\exp\left[-j\omega\left(\frac{f_s d\cos\theta}{c}+l\right)\right] \tag{7}$$

and ω is the frequency in radians, θ is the direction of arrival, c is the speed of sound in air, f_s is the sampling frequency, d_n is the distance of the nth microphone from the origin, and $x_{n,l}$ is the lth coefficient of the nth FIR filter. If θ_d is the desired steering angle of the beamformer, the WNG of the beamformer is given by [3]

$$G_W(\mathbf{x},\mathbf{d},\omega)=\frac{|\mathbf{g}(\mathbf{d},\omega,\theta_d)^T\mathbf{x}|^2}{\|\mathbf{A}(\omega)\mathbf{x}\|_2^2} \tag{8}$$

where

$$\mathbf{A}(\omega)=\mathbf{I}_N\otimes\mathbf{a}(\omega)^T$$
$$\mathbf{a}(\omega)^T=\left[1\,e^{-j\omega}\,\cdots\,e^{-j(L-1)\omega}\right]^T \tag{9}$$

and $\mathbf{I}_N$ is an $N\times N$ identity matrix, $\otimes$ is the Kronecker product, and $\|\mathbf{v}\|_2$ is the L_2 norm of vector $\mathbf{v}$.

2.1. Passband error

If $B_d(\omega,\theta)$ is the desired beampattern at a certain frequency and direction, the squared magnitude error between the beamformer response and the desired beampattern is given by

$$e_b(\mathbf{z},\omega,\theta)=\left|B(\mathbf{x},\mathbf{d},\omega,\theta)\right|^2-\left|B_d(\omega,\theta)\right|^2 \tag{10}$$

where

$$\mathbf{z}^T=\left[\mathbf{x}^T\,\mathbf{d}^T\right] \tag{11}$$

If $\mathbf{z}_k$ is the value of $\mathbf{z}$ at the start of the kth iteration and $\boldsymbol{\delta}$ is the update to $\mathbf{z}_k$, the updated value of the squared magnitude error can be estimated by a linear approximation

$$e_b(\mathbf{z}_k+\boldsymbol{\delta},\omega,\theta)\approx e_b(\mathbf{z}_k,\omega,\theta)+\nabla e_b(\mathbf{z}_k,\omega,\theta)^T\boldsymbol{\delta} \tag{12}$$

which becomes more accurate as $\|\boldsymbol{\delta}\|_2$ gets smaller.

The L_p-norm of the passband squared magnitude error for the kth iteration is given by [3]

$$\begin{aligned}\mathcal{E}_p^{(pb)}(\mathbf{z}_k)&=\left[\int_{\Omega}\int_{\Theta_{pb}}\left|e_b(\mathbf{z}_{k+1},\omega,\theta)\right|^p d\theta d\omega\right]^{1/p}\\&\approx\kappa_{pb}\left[\sum_{r=1}^{R}\sum_{s=1}^{S_{pb}}\left|e_b(\mathbf{z}_{k+1},\omega_r,\theta_s)\right|^p\right]^{1/p}\\&\approx\left[\sum_{r=1}^{R}\sum_{s=1}^{S_{pb}}\left|\kappa_{pb}e_b(\mathbf{z}_k,\omega_r,\theta_s)\right.\right.\\&\qquad\left.\left.+\kappa_{pb}\nabla e_b(\mathbf{z}_k,\omega_r,\theta_s)^T\boldsymbol{\delta}\right|^p\right]^{1/p},\quad\omega_r\in\Omega,\ \theta_s\in\Theta_{pb}\end{aligned} \tag{13}$$

where Ω is the frequency band of interest, $\Theta_{pb}=[\theta_{pl},\theta_{ph}]$ is the angular passband, and κ_{pb} is a constant. Expressing (13) in matrix form we get