Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Design of time-domain modal beamformer for broadband spherical microphone arrays

Shefeng Yan and Chaohuan Hou

Institute of Acoustics, Chinese Academy of Sciences, Beijing, China

PACS: 43.60.Fg Acoustic array systems and processing, beam-forming

ABSTRACT

An approach to real-valued time-domain implementation of modal beamformer for spherical microphone arrays is proposed. The advantage of the time-domain implementation is that we can update the beamformer when each new snapshot arrives. Our technique is based on a modified filter-and-sum spherical harmonics domain (SHD) beamforming structure. The time series received at the microphones are converted into SHD data using spherical Fourier transform. The SHD data input to the steering unit and then feed a bank of finite impulse response (FIR) filters. The filter outputs are summed to produce the beamformer output time series. The FIR filters tap weights are optimally designed by making a compromise among multiple conflicting array performance measures such as directivity, mainlobe spatial response variation (MSRV), sidelobe level, and robustness. The design problem is formulated as a multiply constrained problem which is solved using second-order cone programming (SOCP). Results of simulations show good performance of the proposed time-domain SHD beamformer design approach.

INTRODUCTION

Spherical Harmonics Domain (SHD) beamforming technology has recently become an important research issue in three-dimensional (3D) sound reception, sound field analysis for room acoustics, direction of arrival (DOA) estimation, and so on. A spherical array is more flexible than other array geometries for forming 3D beampattens, and the modal beamforming can be performed using the elegant spherical harmonics framework. Several modal beamforming approaches to spherical arrays have been studied, e.g., regular phase-mode pattern design [1], non-adaptive and adaptive multiple-null steering techniques [2], Dolph-Chebyshev pattern design approach [3], and optimal beamforming methods [4-6], etc.

The studies presented above, however, are all based on a signal model in the frequency domain, where complex-valued modal transformation and array processing is employed. In order to achieve a broadband beamformer, which is usually required for speech and audio applications, the broadband array signals are decomposed into narrow frequency bins using the discrete Fourier transform (DFT) and each frequency bin is independently processed using a narrowband beamforming algorithm, and then an inverse DFT is employed to synthesize the broadband output signal. Since the frequency-domain implementation is performed with block processing, it might be unsuitable for time-critical speech and audio applications due to its associated time delay.

It is well known that, in classical element space array processing, the broadband beamformer can be implemented in the time domain using the filter-and-sum structure [7]. The key point of the filter-and-sum beamformer design is how to calculate the FIR filters' tap weights, in order to achieve the desired beamforming performance.

The spherical array modal beamforming can also be implemented in the time domain with the real-valued modal transformation and the filter-and-sum beamforming structure. Meyer and Elko recently proposed a novel time-domain implementation structure for a spherical array modal beamformer [8], within the spherical harmonics framework. The real and imaginary parts of the spherical harmonics are employed as the spherical Fourier transform basis to convert the time domain broadband signals to the real-valued spherical harmonics domain, and the look direction of the beamformer can be tactfully decoupled from its beampattern shape. To achieve a frequency independent beampattern, Meyer and Elko proposed to employ inverse filters to decouple the frequency-dependent components in each signal channel, however, such kind of inverse filtering could damage the system robustness [1]. Moreover, since no systematic performance analysis framework has been formulated, all the mutually conflicting broadband beamforming performance measures, such as directivity factor, sidelobe level, and robustness, etc. cannot be effectively controlled.

In this paper, an optimal broadband modal beamforming framework implemented in the time domain is presented. Our technique is based on a modified filter-and-sum modal beamforming structure. We derive the expression for the array response, the beamformer output power against both isotropic noise and spatially white noise, and the mainlobe spatial response variation (MSRV) in terms of the FIR filters' tap weights. With the aim of achieving a suitable trade-off among multiple conflicting performance measures (e.g., directivity index, robustness, sidelobe level, mainlobe response variation, etc.), we formulate the FIR filters' tap weights design problem to a multiply constrained optimization problem which is computationally tractable.

BACKGROUND

Spherical Fourier transform

The standard Cartesian (x,y,z) and spherical (r,θ,ϕ) coordinate systems are used. Consider a unit magnitude plane wave impinging on a sphere of radius a from direction $\Omega_0=(\theta_0,\phi_0)$. The spherical harmonics domain expression of the sound pressure on the sphere surface at an observation point $\Omega_s=(\theta_s,\phi_s)$ is given by [9]

$$p_{nm}(ka,\Omega_0)=b_n(ka)[Y_n^m(\Omega_0)]^* . \quad (1)$$

where $k=\omega/c$ is the wavenumber with c being the sound speed, and $\omega=2\pi f$ being the temporal radian frequency with f being the frequency, the superscript * denotes complex conjugation, $b_n(ka)$ is a function of array configuration, with available analytical expressions [9], and Y_n^m is the spherical harmonics of order n and degree m given by $Y_n^m(\Omega)=\sqrt{[(2n+1)(n-m)!]/[4\pi(n+m)!]}P_n^m(\cos\theta)e^{im\phi}$, where $P_n^m(\cos\theta)$ denotes the associated Legendre function and $i=\sqrt{-1}$.

The sound pressure is spatially sampled at the microphone positions Ω_s , $s=1,\cdots,M$, where M is the number of microphones. The microphone positions are required to satisfy the following discrete orthonormality condition:

$$\sum_{s=1}^{M}\alpha_s Y_{n'}^{m'}(\Omega_s)[Y_n^m(\Omega_s)]^*=\delta_{n-n'}\delta_{m-m'} , \quad (2)$$

where $\delta_{n-n'}$ and $\delta_{m-m'}$ are the Kronecker delta functions, and α_s is a real value depending on sampling scheme. For uniform sampling, which is assumed through this paper, in order that $\sum_{s=1}^{M}\alpha_s=\int_{\Omega\in S^2}d\Omega=4\pi$, we have $\alpha_s\equiv 4\pi/M$.

In order to compute up to N th order spherical harmonics, the number of microphones M should be larger than or equal to $(N+1)^2$ to avoid spatial aliasing.

We assume that the time series received at the sth microphone is $x_s(t)$ and the frequency-domain notation is $x(f,\Omega_s)$. Its discrete spherical Fourier transform is given by

$$x_{nm}(f)=\sum_{s=1}^{M}\alpha_s x(f,\Omega_s)[Y_n^m(\Omega_s)]^* . \quad (3)$$

Using (3), the sound field is transformed from the spatial (element-space) domain into the spherical harmonics domain.

Spherical harmonics domain beamforming

Using spherical harmonics domain beamforming, the array output, denoted by $y(f)$, can be calculated as [4]:

$$y(f)=\sum_{n=0}^{N}\sum_{m=-n}^{n}x_{nm}(f)w_{nm}^*(f) , \quad (4)$$

where $w_{nm}^*(f)$ are the weight function. In vector notation,

$$y(f)=\mathbf{w}_b^H(f)\mathbf{x}_b(f) , \quad (5)$$

where $\mathbf{x}_b=\mathrm{vec}(\{[x_{nm}]_{m=-n}^{n}\}_{n=0}^{N})=[x_{00},\cdots,x_{nm},\cdots,x_{NN}]^T$ and $\mathbf{w}_b=\mathrm{vec}(\{[w_{nm}]_{m=-n}^{n}\}_{n=0}^{N})$ with $\mathrm{vec}(\cdot)$ denoting stacking all the entries in the parentheses to obtain an $(N+1)^2\times 1$ column vector. $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and the Hermitian transpose, respectively.

The array output power is given by

$$P_{out}(f)=\mathbf{w}_b^H(f)\mathbf{R}_b(f)\mathbf{w}_b(f) , \quad (6)$$

where $\mathbf{R}_b(f)$ is the covariance matrix of $\mathbf{x}_b$.

A modal beamformer that has a rotational symmetrical beampattern around the look direction Ω_0 can be obtained as long as the array weights take the form [1]

$$w_{nm}^*(f)=\sqrt{4\pi/(2n+1)}c_n(f)Y_n^m(\Omega_0) . \quad (7)$$

which means the weights are divided into two parts, of which $\sqrt{4\pi/(2n+1)}Y_n^m(\Omega_0)$ act as the steering units that are responsible for steering the look direction by Ω_0 and $c_n(f)$ act as pattern generation.

The robustness is an important measure of array performance and is commonly quantified by the white noise gain (WNG), i.e., array gain against white noise. Assuming that $\alpha_s\cong 4\pi/M$, WNG is given by

$$WNG(f)\cong\frac{4\pi/M}{\sum_{n=0}^{N}c_n^*(f)c_n(f)}=\frac{4\pi/M}{\mathbf{c}^H(f)\mathbf{c}(f)} , \quad (8)$$

where $\mathbf{c}=[c_0,\cdots,c_n,\cdots,c_N]^T$.

The directivity pattern, denoted by $B(f,\Omega)$, is given by

$$\begin{aligned}B(f,\Omega)&=\sum_{n=0}^{N}\sum_{m=-n}^{n}p_{nm}(ka,\Omega)w_{nm}^*(f)\\&=\sum_{n=0}^{N}c_n(f)b_n(ka)\sqrt{\frac{2n+1}{4\pi}}P_n(\cos\Theta) , \quad (9)\end{aligned}$$

where P_n is the Legendre polynomial and Θ is the angle between Ω and Ω_0 .

Using (7) in (4) gives

$$y(f)=\sum_{n=0}^{N}\left[\sqrt{\frac{4\pi}{2n+1}}\sum_{m=-n}^{n}x_{nm}(f)Y_n^m(\Omega_0)\right]c_n(f) . \quad (10)$$

TIME-DOMAIN IMPLEMENTATION STRUCTURE

The mathematical analysis of the modal transformation and beamforming was discussed above for complex spherical harmonics. We next consider the time-domain implementation of the broadband modal beamformer. Since the real-valued coefficients are more suitable for a time-domain implementation, we can work with the real and imaginary parts of the spherical harmonics domain data.

We assume that the sampled broadband time series received at the sth microphone is $x_s(l)=x_s(t)|_{t=lT_s}$, where T_s is the sampling interval. Considering that $Y_n^m(\Omega)$ is independent of frequency, similar to (3), the broadband spherical harmonics domain data is given

$$x_{nm}(l)=\sum_{s=1}^{M}\alpha_s x_s(l)[Y_n^m(\Omega_s)]^*,\quad l=1,2,\cdots, \tag{11}$$

where $x_{nm}(l)$ is the time-domain representation of $x_{nm}(f)$ in (3).

We can apply the filter-and-sum structure to a modal beamformer. That is, we place a bank of real-valued FIR filters at the output of the steering unit, the filters act as the role of complex weighting $c_n(f)$ in a broadband frequency band. Let $\mathbf{h}_n$ be the impulse response of the FIR filter corresponding to the spherical harmonics of order n, i.e., $\mathbf{h}_n=[h_{n1},h_{n2},\cdots,h_{nL}]^T$, $n=0,\cdots,N$. Here, L is the length of each FIR filter. Performing the inverse Fourier transform to (10) and considering that the response of the filter $\mathbf{h}_n$ over the working frequency band is approximately equal to $c_n(f)$, the time-domain beamformer output, denoted by $y(l)$, can be given by

$$\begin{aligned} y(l) &= \sum_{n=0}^{N}\left\{\left[\sqrt{\frac{4\pi}{2n+1}}\sum_{m=-n}^{n}\left(\sum_{s=1}^{M}\alpha_s x_s(l)[Y_n^m(\Omega_s)]^*\right)Y_n^m(\Omega_0)\right]*\mathbf{h}_n\right\} \\ &= \sum_{n=0}^{N}\left\{x_n(l,\Omega_0)*\mathbf{h}_n\right\}. \end{aligned} \tag{12}$$

where $*$ denotes the convolution and

$$\begin{aligned} x_n(l,\Omega_0) &= \sqrt{\frac{4\pi}{2n+1}}\sum_{m=-n}^{n}\left(\sum_{s=1}^{M}\alpha_s x_s(l)[Y_n^m(\Omega_s)]^*\right)Y_n^m(\Omega_0) \\ &= \sqrt{\frac{4\pi}{2n+1}}\left\{\tilde{x}_{n0}(l)Y_n^0(\Omega_0)+2\sum_{m=1}^{n}\tilde{x}_{nm}(l)]\,\mathrm{Re}[Y_n^m(\Omega_0)]\right. \\ &\quad \left.+2\sum_{m=1}^{n}\breve{x}_{nm}(l)\,\mathrm{Im}[Y_n^m(\Omega_0)]\right\}, \end{aligned} \tag{13}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote the real part and imaginary part, respectively, $\tilde{x}_{nm}(l)=\sum_{s=1}^{M}\alpha_s x_s(l)\,\mathrm{Re}[Y_n^m(\Omega_s)]$ and $\breve{x}_{nm}(l)=\sum_{s=1}^{M}\alpha_s x_s(l)\,\mathrm{Im}[Y_n^m(\Omega_s)]$. Eq.(13) can be rewritten as

$$\begin{aligned} x_n(l,\Omega_0) &= \tilde{x}_{n0}(l)P_n^0(\cos\theta_0)+2\sum_{m=1}^{n}\left\{\sqrt{\frac{(n-m)!}{(n+m)!}}\cdot\right. \\ &\quad \left. P_n^m(\cos\theta_0)[\tilde{x}_{nm}(l)\cos(m\phi_0)+\breve{x}_{nm}(l)\sin(m\phi_0)]\right\}. \end{aligned} \tag{14}$$

According to (12) and (14), the time-domain implementation of the broadband modal beamformer is given in Figure. 1.

In this beamformer implementation structure, the modal transformation unit and steering unit is similar as that suggested by Elko et al [8]. However, the parameters here are different from that in [8]. The FIR pattern generation unit is the contribution of this paper. Our goal is then to choose the impulse responses of these FIR filters to achieve the desired frequency-wavenumber response of the modal beamformer.

The total weighting function in the pattern generation unit corresponding to the nth order spherical harmonics at frequency f is given by

$$\hat{c}_n(f)=e^{-j2\pi f T_0}\mathbf{h}_n^T\mathbf{e}(f),\quad n=0,1,\cdots,N. \tag{15}$$

where $\mathbf{e}(f)=[1,e^{-j2\pi f T_s},\cdots,e^{-j(L-1)2\pi f T_s}]^T$ and $T_0=-(L-1)T_s/2$ is introduced to compensate the inherent group delay of the FIR filters.

We use $\hat{c}_n(k)$ in (15) in lieu of $c_n(k)$ in (9) and let $\eta=e^{-j2\pi f T_0}$ to obtain

$$B(f,\Omega)=\sum_{n=0}^{N}b_n(ka)\sqrt{\frac{2n+1}{4\pi}}P_n(\cos\Theta)\eta\mathbf{h}_n^T\mathbf{e}(f). \tag{16}$$

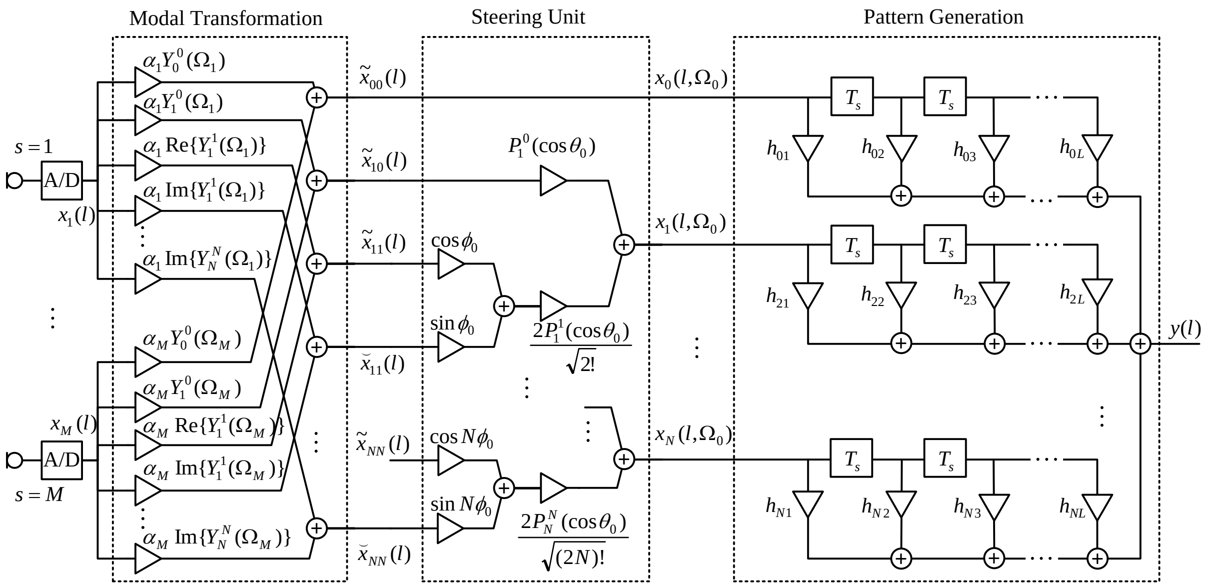

Figure 1. Time-domain implementation of broadband modal beamformer.

Let $a_n(f,\Theta)=b_n(ka)\sqrt{\frac{2n+1}{4\pi}}P_n(\cos\Theta)\eta$, $\mathbf{a}=[a_0,\cdots,a_n,\cdots,a_N]^T$, and define an $(N+1)L\times 1$ composite vector $\mathbf{h}=[\mathbf{h}_0^T,\mathbf{h}_1^T,\cdots,\mathbf{h}_N^T]^T$. Eq.(16) can then be rewritten as

$$\begin{aligned}B(f,\Omega) &= \sum_{n=0}^{N} a_n(f,\Theta)\mathbf{h}_n^T\mathbf{e}(f)\\ &=[\mathbf{a}(f,\Theta)\otimes\mathbf{e}(f)]^T\mathbf{h}\\ &=\mathbf{u}^T(f,\Theta)\mathbf{h} = \mathbf{h}^T\mathbf{u}(f,\Theta)\,, \qquad (17)\end{aligned}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{u}(f,\Theta)=\mathbf{a}(f,\Theta)\otimes\mathbf{e}(f)$.

The distortionless constraint in the spherical harmonics domain is [4]

$$\mathbf{h}^T\mathbf{u}(f,0)=4\pi/M\ . \qquad (18)$$

DESIGN OF FIR FILTERS

The isotropic noise covariance matrix is given by

$$\mathbf{Q}_{biso}(f)=\frac{\sigma_n^2(f)}{4\pi}\mathrm{diag}\{\mathbf{b}_b(ka)\circ\mathbf{b}_b^*(ka)\}\,, \qquad (19)$$

where $\sigma_n^2(f)$ is the power spectral density of the isotropic noise, $\mathbf{b}_b=\mathrm{vec}(\{[b_n]_{m=-n}^{n}\}_{n=0}^{N})$, $\circ$ denotes the Hadamard (i.e., element-wise) product of two vectors, and $\mathrm{diag}\{\cdot\}$ denotes a square matrix with the elements of its arguments on the diagonal.

Consider a special case with only isotropic noise impinging on the microphone array. We use (6) with $\mathbf{R}_b(f)$ replaced by the isotropic noise covariance matrix $\mathbf{Q}_{biso}(f)$ to obtain the isotropic noise-only beamformer output power

$$\begin{aligned}P_{isoout}(f) &= \mathbf{w}_b^H(f)\mathbf{Q}_{biso}(f)\mathbf{w}_b(f)\\ &=\mathbf{c}^T(f)\mathbf{Q}_{ciso}(f)\mathbf{c}^*(f)\\ &=\mathbf{h}^T\mathbf{Q}_{hiso}(f)\mathbf{h}\,, \qquad (20)\end{aligned}$$

where $\mathbf{Q}_{ciso}(f) = \sigma_n^2(f)/(4\pi)\cdot\mathrm{diag}\{\mathbf{b}_c(ka)\circ\mathbf{b}_c^*(ka)\}$ with $\mathbf{b}_c(ka)=[b_0(ka),b_1(ka),b_2(ka),\cdots,b_N(ka)]^T$ and

$$\mathbf{Q}_{hiso}(f)=[\mathbf{I}_{(N+1)\times(N+1)}\otimes\mathbf{e}(f)]\mathbf{Q}_{ciso}(f)[\mathbf{I}_{(N+1)\times(N+1)}\otimes\mathbf{e}(f)]^H\ .$$

For a broadband isotropic noise that occupies the frequency band $[f_L,f_U]$ with f_L and f_U being respectively the lower and upper bound frequency, its broadband covariance matrix, denoted by $\overline{\mathbf{Q}}_{hiso}$, can be given by performing the integration with respect to f over the region $[f_L,f_U]$

$$\overline{\mathbf{Q}}_{hiso}=\int_{f_L}^{f_U}\mathbf{Q}_{hiso}(f)\ . \qquad (21)$$

where the integration can be approximated by performing summation.

Assume that the isotropic noise has a flat spectrum $\sigma_n^2(f)=1$ over the frequency band $[f_L,f_U]$. The broadband isotropic noise-only beamformer output power is

$$\overline{P}_{isoout}=\mathbf{h}^T\overline{\mathbf{Q}}_{hiso}\mathbf{h}\ . \qquad (22)$$

Consider another special case where only spatially white noise with power spectral density $\sigma_{nw}^2(f)$ impinging on the microphone array. In the case of $\alpha_s\cong 4\pi/M$, the spatially white noise-only beamformer output power, denoted by $P_{wout}(f)$, is given by

$$\begin{aligned}P_{wout}(f) &\cong \frac{4\pi\sigma_{nw}^2(f)}{M}\sum_{n=0}^{N}\sum_{m=-n}^{n}|w_{nm}(f)|^2\\ &=\frac{4\pi\sigma_{nw}^2(f)}{M}\sum_{n=0}^{N}|\mathbf{h}_n^T\mathbf{e}(f)|^2\ . \qquad (23)\end{aligned}$$

Assume that the spatially white noise has a flat spectrum $\sigma_{nw}^2(f)=1$ over the whole frequency band $[0,f_s/2]$. The broadband beamformer output power, denoted by $\overline{P}_{wout}$, is given by

$$\overline{P}_{wout}=\int_0^{f_s/2}P_{wout}(f) = \frac{4\pi}{M}\sum_{n=0}^{N}\mathbf{h}_n^T\mathbf{h}_n = \frac{4\pi}{M}\mathbf{h}^T\mathbf{h}\ . \qquad (24)$$

The broadband white noise gain, denoted by $BWNG$, is then defined as

$$BWNG=\frac{(4\pi/M)^2}{\overline{P}_{wout}}=\frac{4\pi/M}{\mathbf{h}^T\mathbf{h}}\ . \qquad (25)$$

A common measure of performance of an array is the directivity. The directivity factor $D(f)$, or directive gain, can be interpreted as the array gain against isotropic noise, which is given by

$$D(f)=\frac{\sigma_n^2(f)(4\pi/M)^2}{\mathbf{h}^T\mathbf{Q}_{hiso}(f)\mathbf{h}}\ . \qquad (26)$$

Frequently, we express the directivity factor in dB and refer to it as the directivity index (DI), $DI(f)=10\log_{10}D(f)$.

The MSRV is defined as

$$\gamma_{MSRV}(f,\Theta) = |\mathbf{h}^T\mathbf{u}(f,\Theta)-\mathbf{h}^T\mathbf{u}(f_0,\Theta)|\,, \qquad (27)$$

where f_0 is a chosen reference frequency.

Let $f_k\in[f_L,f_U]$ $(k=1,2,\cdots,K)$, $\Theta_j\in\mathbf{\Theta}_{ML}$ $(j=1,\cdots,N_{ML})$, and $\Theta_i\in\mathbf{\Theta}_{SL}$ $(i=1,\cdots,N_{SL})$ be a chosen (uniform or non-uniform) grid that approximates the frequency band $[f_L,f_U]$, the mainlobe region $\mathbf{\Theta}_{ML}$, and the sidelobe region $\mathbf{\Theta}_{SL}$, respectively. We define an $N_{ML}K\times 1$ column vector $\boldsymbol{\gamma}_{MSRV}$ and an $N_{SL}K\times 1$ column vector $\mathbf{B}_{SL}$, whose entries are respectively given by

$$[\boldsymbol{\gamma}_{MSRV}]_{k+(j-1)K}=\gamma_{MSRV}(f_k,\Theta_j)\,, \qquad (28)$$

$$[\mathbf{B}_{SL}]_{k+(i-1)K}=B(f_k,\Theta_i)\ . \qquad (29)$$

Then, the norm of $\boldsymbol{\gamma}_{MSRV}$, i.e., $\|\boldsymbol{\gamma}_{MSRV}\|_q$, can be used as a measure of the frequency- invariant approximation of the synthesized broadband beampattern over frequencies. The subscript $q\in\{2,\infty\}$ stands for the l_2 (Euclidean) and

l_∞ (Chebyshev) norm, respectively. Similarly, $\|\mathbf{B}_{SL}\|_q$ is a measure of sidelobe behavior.

There are many performance measures by which one may assess the capabilities of a beamformer. Commonly used array performance measures are directivity, MSRV, sidelobe level, and robustness. The trade-off among these conflicting performance measures represents the beamformer design optimization problem. After formulating the broadband spherical harmonics domain beampattern $B(f,\Omega)$ (17), the broadband isotropic noise-only beamformer output power $\overline{P}_{isoout}$ (22), the broadband white noise gain $BWNG$ (25), the mainlobe spatial response variation vector $\boldsymbol{\gamma}_{MSRV}$ (28), and the sidelobe behavior vector $\mathbf{B}_{SL}$ (29), the optimal array pattern synthesis problem for broadband modal beamformer can be formulated as

$$
\begin{aligned}
&\min_{\mathbf{h}} \mu_\ell\,, \quad \ell=\{1,2,3,4\}\,,\\
&\text{subject to } \; B(f_k,\Omega_0)=4\pi/M\,, \quad k=1,2,\cdots,K\\
&\qquad \overline{P}_{isoout}\le \mu_1\,, \quad \|\boldsymbol{\gamma}_{MSRV}\|_{q_1}\le \mu_2\,,\\
&\qquad \|\mathbf{B}_{SL}\|_{q_2}\le \mu_3\,, \quad BWNG^{-1}\le \mu_4\,, \qquad (30)
\end{aligned}
$$

where $q_1,q_2\in\{2,\infty\}$, and $\{\mu_\ell\}_{\ell=1}^4$ include a cost function and three user parameters. Using the similar process technique as proposed in our earlier paper [7] for classical array processing, the optimization problem (30) can be reformulated in a convex form as the so-called SOCP which can be solved efficiently using an SOCP solver such as SeDuMi [10].

SIMULATION EXAMPLES

In this section, a numerical example is provided to illustrate the performance of the proposed approach.

We consider a rigid spherical array of radius 4.2 cm with $M=32$ microphones located at the center of the faces of a truncated icosahedron. An order of $N=4$ is used for sound field decomposition and $\alpha_s\equiv 4\pi/M$. The sampling frequency is $f_s=14700$ Hz. The frequency band [500 Hz, 5000 Hz] is discretized using $K=51$ frequency grids $f_k=f_L\cdot 10^{\lg(f_U/f_L)*(k-1)/(K-1)}$, $k=1,2,\cdots,K$. The length of the FIR filters is $L=65$. Unless otherwise stated, we assume $\boldsymbol{\Theta}_{ML}=[0°:2°:40°]$ and $\boldsymbol{\Theta}_{SL}=[48°:2°:180°]$, which means a uniform grid of 2° is used to discretize the directions.

Assume that we want to design a Time-Domain Robust Maximal-directivity (TDRMD) modal beamformer. The optimization problem (30) in this case can be formulated as $\ell=1$, $\mu_2=\infty$, $\mu_3=\infty$, and μ_4 being a user parameter.

Assume that $\mu_4=4\pi/M$. The FIR filter $\mathbf{h}$ is determined by solving the optimization problem and its subvectors $\mathbf{h}_0,\mathbf{h}_1,\cdots,\mathbf{h}_N$ are shown in Figure 2(a). We substitute $\mathbf{h}$ into (15) to get $\hat{c}_n(f)$ and display them in Figure 2(b). Using (17), the beampattern as a function of frequency and angle are calculated on a grid of points in frequency and angle. The resulting beampatterns are shown in Figure 2(c), where we have included a normalization factor $M/4\pi$ so the amplitudes of the patterns at the look direction are equal to unity (or to 0 dB). The DI and WNG of the time-domain modal beamformer are calculated by using (26) and (8), respectively. The results are shown in Figure 2(d) for various frequencies.

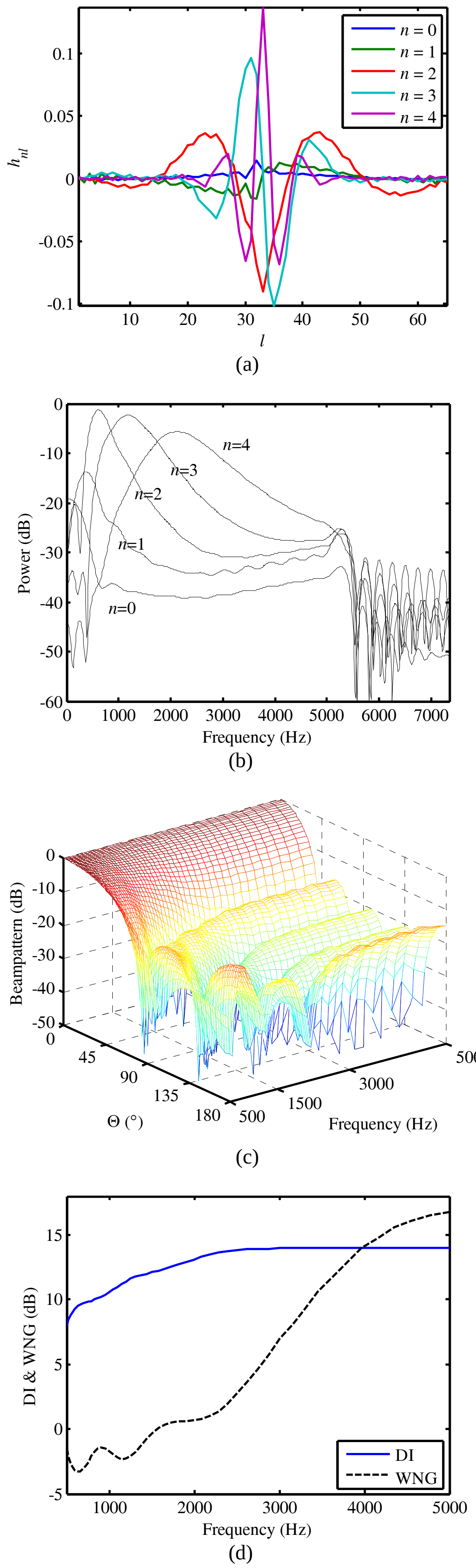

Figure 2. Performance of beamformer using robust maximal directivity design. (a) The 5 FIR filters, (b) the weighting function, (c) the beampattern, and (d) the DI and WNG at various frequencies.

It is seen from Figure 2(d) that the WNG of this beamformer is higher than -3 dB, which is suitable for most applications. The DI of this beamformer is also good. The results show that this design can provide a good tradeoff between the directivity and the robustness.

CONCLUSION

The real-valued time-domain implementation of the broadband modal beamformer in the spherical harmonics domain has been presented. The broadband modal beamformer is composed of the modal transformation unit, the steering unit, and the pattern generation unit. The pattern generation unit is independent of the steering direction and is implemented using a filter-and-sum structure. The elegant spherical harmonics framework leads to more computationally efficient optimization algorithm and implementation scheme than conventional element-space based approaches. The broadband array response, the beamformer output power against both isotropic noise and spatially white noise, and the mainlobe spatial response variation are all expressed as the functions of the FIR filters' tap weights. The FIR filters design problem is formulated as a multiply constrained problem, which ensures that the resulting beamformer can provide a suitable trade-off among multiple conflicting array performance measures such as directivity, mainlobe spatial response variation, sidelobe level, and robustness. The performance of the proposed approach is demonstrated by a number of simulations.

ACKNOWLEDGE

The author gratefully acknowledges the support of K. C. Wong Education Foundation, Hong Kong.

REFERENCES

1 J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. ICASSP*, **2**, 1781-1784 (2002)
2 B. Rafaely, "Spherical microphone array with multiple nulls for analysis of directional room impulse responses," in *Proc. ICASSP*, 281-284 (2008)
3 A. Koretz and B. Rafaely, "Dolph-Chebyshev beampattern design for spherical arrays," *IEEE Trans. Signal Process.* **57**, 2417-2420 (2009)
4 S. Yan, H. Sun, U. P. Svensson, X. Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Trans. Audio Speech Lang. Process*. to appear (2010)
5 Z. Y. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming," *IEEE Trans. Audio Speech Lang. Process.* **15**, 702-714 (2007)
6 H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition.* (Springer-Verlag, Berlin Heidelberg, 2007)
7 S. Yan, Y. Ma, and C. Hou, "Optimal array pattern synthesis for broadband arrays," *J. Acoust. Soc. Am.*, **122**, 2686-2696 (Nov. 2007)
8 G. W. Elko, R. A. Kubli, and J. Meyer, "Audio system based on at least second-order eigenbeam," *United States Patent* WO 03/061336 A1 (2003)
9 E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography.* (Academic Press, New York, 1999)
10 J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Software*, **11-12**, 625-653 (1999).