Three-dimensional wave-domain acoustic contrast control using a circular loudspeaker array

Zerui Han,[1,2] **Ming Wu,**[1,2] **Qiaoxi Zhu,**[3] **and Jun Yang**[1,2,a)]

[1)] *Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China*

[2)] *School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China*

[3)] *Centre for Audio, Acoustics and Vibration, Faculty of Engineering and IT, University of Technology Sydney, NSW 2007, Australia*

hanzerui@mail.ioa.ac.cn,

mingwu@mail.ioa.ac.cn,

qiaoxi.zhu@gmail.com,

jyang@mail.ioa.ac.cn

(Dated: 16 April 2019)

Abstract: This paper proposes a three-dimensional wave-domain acoustic contrast control method to reproduce a multizone sound field using a circular loudspeaker array. In this method, sound field analysis is based on spherical harmonic decomposition, and the loudspeaker weights are obtained by maximizing the acoustic energy contrast between the predefined bright zone and dark zone. Simulation results show that the proposed method provides good multizone separation performance over a large spatial region and requires lower-order spherical harmonics, resulting in a much lower number of microphones required to measure the acoustic transfer functions.

[a] Author to whom correspondence should be addressed.

1. Introduction

Multizone sound field reproduction aims to generate individual sound fields within different spatial zones by a loudspeaker array. Acoustic contrast control (ACC) is a common method for achieving multizone sound control by maximizing the contrast of the acoustic energy between the target zones[1]. The conventional ACC method is implemented at a set of discrete sampling points over the entire region of interest. This method requires numerous sampling points to measure the acoustic transfer functions (ATFs) when broadband reproduction over large target zones is desired.

Recently, the wave-domain method, which optimizes a sound field over a region by controlling the expansion coefficients, has attracted increasing attention in the research of sound field control[2,3]. In our previous work, we introduced a wave-domain acoustic contrast control (WDACC) method based on circular harmonic expansion to manipulate a two-dimensional (2D) multizone sound field[4]. Using the same number of microphones, the WDACC method outperforms the conventional ACC method in terms of acoustic contrast and array gain over the 2D regions.

In this paper, we extend the idea of WDACC to three-dimensional (3D) multizone sound field control, with improved control in the vertical direction at the sound zones. In this method, spherical harmonic decomposition is used to represent the 3D sound field and formulate the acoustic energy over the region of interest. Instead of a spherical array, a circular loudspeaker array in the horizontal plane is employed to reproduce the 3D multizone sound field because a circular array is preferred in practical realization and can be used for application scenarios, e.g., smart speakers. The simulation results show that compared with conventional ACC method, the proposed method has better acoustic contrast performance at high frequencies in both free-field and reverberation conditions. Furthermore, this method achieves good multizone effects with lower truncation orders, indicating a significant reduction in the number of microphones required to measure the ATFs. The remainder of this paper is organized as follows. Section 2 presents a spherical harmonic description of the first-order loudspeakers and the acoustic potential energy density. Section 3 provides the mathematical formulation of the three-dimensional WDACC method. Simulation results are presented in Section 4. Conclusions are given in Section 5.

2. Spherical harmonic description of 3D sound fields

In this section, spherical harmonics are first introduced to describe an interior sound field. We also derive the spherical harmonic expansions of a monopole and a dipole as sound sources and further investigate the acoustic potential energy density in the 3D wave domain.

The solution to the Helmholtz equation can be expressed in spherical coordinates $\mathbf{r} = (r, \theta, \phi)$, where $r = ||\mathbf{r}||$, θ is the elevation angle, and ϕ is the azimuth angle. The internal sound pressure field for a source-free region of space is[5]

$$p(\mathbf{r}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} a_n^m(k) j_n(kr) Y_n^m(\theta, \phi), \tag{1}$$

where $k = 2\pi f/c$ is the wavenumber, f is the frequency, c is the speed of sound in air, $a_n^m(k)$ is a set of sound field coefficients, $j_n(\cdot)$ is the spherical Bessel function of integer order n, and the spherical harmonic is defined as

$$Y_n^m(\theta, \phi) = \mathcal{P}_n^m(\cos\theta)e^{im\phi}, \tag{2}$$

where n and m are the spherical harmonic degree and order, respectively,

$$\mathcal{P}_n^m(\cos\theta) = \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta) \tag{3}$$

is the orthonormalized associated Legendre function[6], and $P_n^m(\cdot)$ is the associated Legendre function. The spherical harmonics are orthogonal over both degree n and order m according to

$$\iint_\Omega Y_n^m(\theta, \phi)^* Y_{n'}^{m'}(\theta, \phi)\mathrm{d}\Omega = \delta_{nn'}\delta_{mm'}, \tag{4}$$

where $\mathrm{d}\Omega = \sin\theta \mathrm{d}\theta\mathrm{d}\phi$ is the differential surface area on the unit sphere and δ_{ij} is the Kronecker delta function.

2.1 Monopole and dipole sources

The acoustic pressure field generated by an ideal monopole located at $\mathbf{r}_s = (r_s, \theta_s, \phi_s)$ in the free field is given by[7]

$$p_m(\mathbf{r}, \mathbf{r}_s, k) = \frac{e^{ik||\mathbf{r}-\mathbf{r}_s||}}{4\pi||\mathbf{r}-\mathbf{r}_s||} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} j_n(kr)Y_n^m(\theta, \phi)h_n(kr_s)Y_n^m(\theta_s, \phi_s)^*, \tag{5}$$

where $h_n(\cdot)$ is the nth order spherical Hankel function of the first kind.

An ideal dipole consists of two point sources opposite in phase and separated by an infinitesimal distance. A dipole located at $\mathbf{r}_s$ and oriented in direction $\mathbf{v}$ creates a pressure[5]

$$p_d(\mathbf{r}, \mathbf{r}_s, k) = \frac{\partial p_m(\mathbf{r}, \mathbf{r}_s, k)}{\partial \mathbf{v}} = -ik\frac{e^{ik||\mathbf{r}-\mathbf{r}_s||}}{4\pi||\mathbf{r}-\mathbf{r}_s||}(1 + \frac{i}{k||\mathbf{r}-\mathbf{r}_s||})\cos\psi, \tag{6}$$

where ψ is the angle between $\mathbf{r} - \mathbf{r}_s$ and $\mathbf{v}$. For a tangential dipole, the orientation $\mathbf{v}$ is the unit vector along the direction of increasing θ_s. To produce a broadband flat frequency response, the dipole sound field can be equalized by dividing by ik[8]. Given that $\partial\mathbf{v} = r_s\partial\theta_s$, the spherical harmonic expansion for an equalized tangential dipole response is derived as

$$\begin{aligned} p_d(\mathbf{r}, \mathbf{r}_s, k) &= \frac{1}{ikr_s}\frac{\partial p_m(\mathbf{r}, \mathbf{r}_s, k)}{\partial\theta_s} \\ &= -\frac{\sin\theta_s}{r_s}\sum_{n=0}^{\infty}\sum_{m=-n}^{n} j_n(kr)Y_n^m(\theta,\phi)h_n(kr_s)\mathcal{P}_n^m(\cos\theta_s)' e^{-im\phi_s}, \end{aligned} \tag{7}$$

where $\mathcal{P}_n^m(\cdot)'$ denotes the first-order derivative of the orthonormalized associated Legendre function. In the following text, the sound pressure generated by the sound source is described as the ATF between the target region and the loudspeaker.

2.2 First-order loudspeaker on a plane

A sound source composed of monopole and dipole components is known as a first-order loudspeaker[8]. The ATF of a first-order loudspeaker consists of a weighted sum of a monopole and an equalized tangential dipole,

$$p(\mathbf{r}, \mathbf{r}_s, k) = \gamma p_m(\mathbf{r}, \mathbf{r}_s, k) + (1-\gamma)p_d(\mathbf{r}, \mathbf{r}_s, k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} b_n^m(\mathbf{r}_s, k)j_n(kr)Y_n^m(\theta,\phi), \tag{8}$$

where $\gamma \in [0, 1]$ is the weighting factor, and

$$b_n^m(\mathbf{r}_s, k) = h_n(kr_s)e^{-im\phi_s}[\gamma\mathcal{P}_n^m(\cos\theta_s) - (1-\gamma)\frac{\sin\theta_s}{r_s}\mathcal{P}_n^m(\cos\theta_s)'].$$

According to the recurrence relation for the associated Legendre functions[9]

$$(1-x^2)\frac{dP_n^m(x)}{dx} = -nxP_n^m(x) + (n+m)P_{n-1}^m(x), \tag{9}$$

we can derive that the orthonormalized associated Legendre functions obey

$$(1-x^2)\mathcal{P}_n^m(x)' = -nx\mathcal{P}_n^m(x) + \sqrt{\frac{(2n+1)(n^2-m^2)}{2n-1}}\mathcal{P}_{n-1}^m(x). \tag{10}$$

Note that in some special cases, it is impossible to reproduce a sound field of some orders using only monopoles or dipoles on a plane. For example, when $\theta_s = \pi/2$, i.e., $\cos\theta_s = 0$, the value of $\mathcal{P}_n^m(0)$ is equal to zero when $n+m$ is an odd integer, and $\mathcal{P}_n^m(0)' = 0$ when $n+m$ is an even integer[3]. This effect is the reason why we utilize first-order loudspeakers to reproduce the 3D sound field.

2.3 Acoustic potential energy density

The acoustic potential energy density in the wave domain over a spherical region $D = \{(r,\theta,\phi) : r \in [0,R], \theta \in [0,\pi], \phi \in [0,2\pi)\}$ is defined as

$$E = \frac{1}{V_D}\iiint_D |p(\mathbf{r},k)|^2 \mathrm{d}V = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} w_n(k,R)|a_n^m(k)|^2, \tag{11}$$

which is derived from the orthogonality property of the spherical harmonics in Eq. (4), where $\mathrm{d}V = r^2\sin\theta\mathrm{d}r\mathrm{d}\theta\mathrm{d}\phi$, $w_n(k,R) = \frac{1}{V_D}\int_0^R |j_n(kr)|^2 r^2 \mathrm{d}r$ is the coefficient weighting function, and $V_D = \frac{4}{3}\pi R^3$ is the volume of the spherical region D.

The energy contribution of each order of the spherical harmonics is controlled by the weighting functions. Figure 1 illustrates a waterfall plot of the power level of $w_n(k,R)$ in decibels (dB) varying with k and n for R set to 0.3 m. As shown in the figure, the weighting functions decrease rapidly after specific orders, indicating that only a part of the spherical harmonics contribute significant energy to the 3D sound field. Thus, the first summation in Eqs. (1), (8) and (11) can be truncated to $N = \lceil ekR/2 \rceil$ terms within a given region of

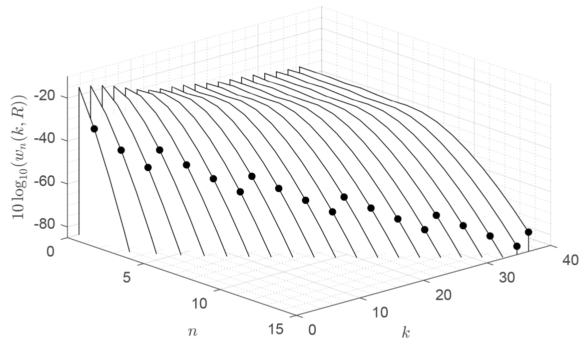


Fig. 1. The power level of $w_n(k, R)$ versus k and n for a specific control region with radius $R = 0.3$ m. The black dots denote the truncated order $N = \lceil ekR/2 \rceil$.

interest[10]. It is sufficient to use $(N + 1)^2$ of the spherical harmonics to represent a sound field within a radius R of interest.

3. 3D wave-domain acoustic contrast control

Consider a circular loudspeaker array of L first-order sources located at $\mathbf{r}_{s,l}, l = 1 \cdots L$. By combining Eqs. (1) and (8), the coefficients of the sound fields reproduced by the loudspeaker array and the ATFs of the first-order loudspeakers are related to the source weights by

$$a_n^m(k) = \sum_{l=1}^{L} b_n^m(\mathbf{r}_{s,l}, k) q_l(k), \tag{12}$$

where $q_l(k)$ is the source weight of the lth loudspeaker. By truncating the sound field to $(N + 1)^2$ terms in Eqs. (1) and (8), this relationship can be rewritten into matrix form as

$$\mathbf{a}(k) = \mathbf{B}(k)\mathbf{q}(k), \tag{13}$$

where $\mathbf{a}(k) = [a_0^0(k), \cdots, a_N^{-N}(k), \cdots, a_N^N(k)]^T$ is the spherical harmonic coefficient vector of the sound field, $\mathbf{q}(k) = [q_1(k), \cdots, q_L(k)]^T$ is the weight vector of the loudspeaker array, and

$$\mathbf{B}(k) = \begin{bmatrix} b_0^0(\mathbf{r}_{s,1}, k) & b_0^0(\mathbf{r}_{s,2}, k) & \dots & b_0^0(\mathbf{r}_{s,L}, k) \\ \vdots & \vdots & \ddots & \vdots \\ b_N^N(\mathbf{r}_{s,1}, k) & b_N^N(\mathbf{r}_{s,2}, k) & \dots & b_N^N(\mathbf{r}_{s,L}, k) \end{bmatrix}$$

is the matrix of the spherical harmonic coefficients of the ATFs between the loudspeaker array and the control zone. For convenience, k is no longer marked in the following text.

The acoustic potential energy density can also be written in the following matrix form by truncating the order of the spherical harmonics as

$$E = \mathbf{a}^H \mathbf{W} \mathbf{a} = \mathbf{q}^H \mathbf{B}^H \mathbf{W} \mathbf{B} \mathbf{q} = \mathbf{q}^H \mathbf{R} \mathbf{q}, \tag{14}$$

where $\mathbf{W} = \mathrm{diag}\{w_0(k, R), w_1(k, R), w_1(k, R), w_1(k, R), \cdots, w_N(k, R)\}$ is a diagonal coefficient weighting matrix, $(\cdot)^H$ represents the Hermitian transpose, and $\mathbf{R} = \mathbf{B}^H \mathbf{W} \mathbf{B}$ denotes the wave-domain correlation matrix. It should be noted that $\mathbf{R}$ and R are different physical quantities.

Next, let there be two disjoint control zones under 3D conditions. Assume that the desired bright zone is a spherical area with radius R_b and that the dark zone is another spherical area with radius R_d. According to Eq. (14), we can define the acoustic potential energy density over the bright zone and the dark zone, E_b and E_d, respectively. Therefore, the reproduced sound field over the regions of interest can be optimized by maximizing the ratio of the contrast among the control zones as

$$\max_{\mathbf{q}} C = \frac{E_b}{E_d} = \frac{\mathbf{q}^H \mathbf{B}_b^H \mathbf{W}_b \mathbf{B}_b \mathbf{q}}{\mathbf{q}^H \mathbf{B}_d^H \mathbf{W}_d \mathbf{B}_d \mathbf{q}} = \frac{\mathbf{q}^H \mathbf{R}_b \mathbf{q}}{\mathbf{q}^H \mathbf{R}_d \mathbf{q}}, \tag{15}$$

where C denotes the acoustic contrast between the bright zone and the dark zone, and the level of this contrast in decibels is defined as $10 \log_{10} C$. The solution to Eq. (15) is given by[1]

$$\mathbf{q} = \Phi\{(\mathbf{R}_d + \delta \mathbf{I})^{-1} \mathbf{R}_b\}, \tag{16}$$

where $\Phi\{\cdot\}$ represents the eigenvector corresponding to the maximum eigenvalue, $\mathbf{I}$ is an identity matrix, and $\delta > 0$ is the Tikhonov regularization parameter.

It is noteworthy that all the spherical harmonic coefficients are with respect to the coordinate system whose origin is at the center of the control region, instead of at the center of the loudspeaker array. This feature makes the proposed method able to avoid the Bessel zeros problem in sound reproduction, meaning that a single-ring circular array can be exploited to produce a 3D sound field, while other methods usually use a multiple cocentered array layout to overcome the Bessel zeros problem[3].

In practical applications, it is generally necessary to precalibrate the ATFs using microphone arrays. For the 2D case, at least $2N+1$ microphones are needed, and for the 3D case, the number of microphones required increases to $(N+1)^2$, usually more in practice, which increases substantially as kR increases[10]. However, Fig. 1 indicates that such a high truncation order is not needed to represent the acoustic energy at high frequencies. Several simulation results are given in the following section to demonstrate this finding.

4. Results and discussions

In the following examples, the speed of sound $c = 344$ m/s and is assumed to be uniform. We use an equiangular circular array of radius $R = 0.1$ m with 6 first-order loudspeakers to implement the simulation. Considering the size of a listener's head, the radii of the spherical bright zone and dark zone are $R_b = R_d = 0.3$ m, and a larger area case where $R_b = R_d = 0.5$ m is also included as a comparison. Both free-field and reverberant conditions are considered. The center of the loudspeaker array is defined as the coordinate origin, and the centers of the bright zone and the dark zone are located at $(1, -1, 1)$ m and $(-1, -1, 0.5)$ m, respectively. The selection of γ should depend on the form of the sound field being controlled according to the source directivities. For example, for vertical sound zones, the dipole component should increase, and for horizontal sound zones, the opposite. In our simulation, we choose $\gamma = 0.2$. The room size is $15 \times 8 \times 5$ m^3, and the reflection coefficients (r_p) are 0.6 and 0.8, respectively. The ATFs in the reverberant room are generated by the image source method[11]. The system setup is illustrated in Fig. 2. To evaluate the acoustic contrast performance over a spatial region, we choose 100 random observation points inside each of the control zones. Multiplicative error of the ATFs is introduced and follows a Gaussian distribution with standard deviations of 3 dB.

We compared WDACC with traditional ACC method in terms of average acoustic contrast over the control region. As mentioned in Sec. 4, the truncation order of the spherical harmonics does not need to be very high. Considering the spherical microphone array in the market[12], the truncation order is limited to 5 and 3, respectively, i.e., $N = \min(\lceil ekR/2 \rceil, 5)$

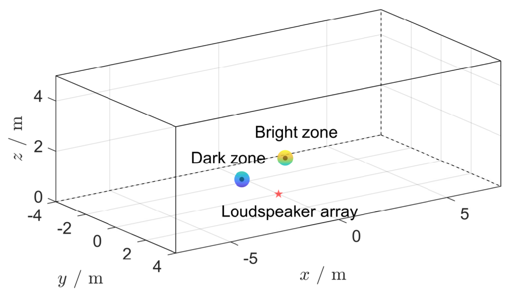
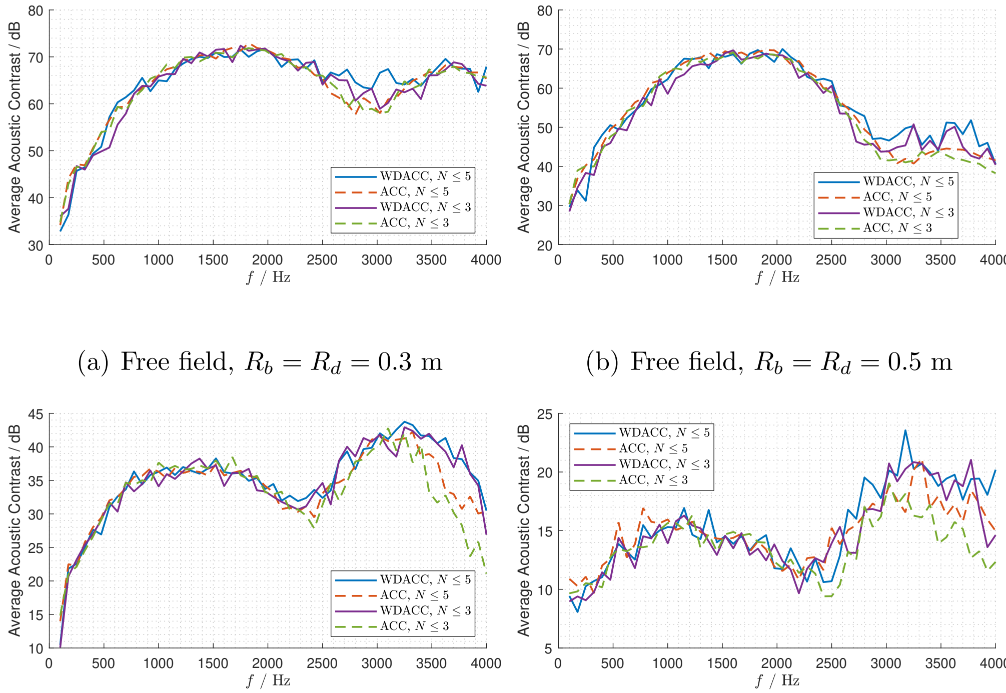


Fig. 2. (Color online) Geometry of the simulation setup. The pentagram denotes the center of the loudspeaker array, and the two black dots are the centers of the control regions. The solid and the dashed lines show the outlines of the room.

and $N = \min(\lceil ekR/2 \rceil, 3)$, while an equiangular open sphere array with around $1.5(N+1)^2$ microphones is employed to measure the ATFs for ACC method. The broadband performance evaluation with varying regularization parameters is considered. δ is varied from 10^{-7} to 10^{-2} at 26 logarithmically spaced values for each frequency over a range of 100 to 4000 Hz. The regularization parameters corresponding to the best average acoustic contrast performance are selected.

Figures 3(a) and 3(b) show the average acoustic contrast levels as functions of the frequency using the proposed WDACC and the conventional ACC method with different sizes of control regions in the free field. As can be seen, when the control regions become larger, both methods degrade at high frequencies. The performances with reduced measurements is close to the full measurement case. The two methods have very similar contrast performance at low frequencies, but WDACC is better at high frequencies, although the superiority is vague. In the free field, measurements around the interest regions are sufficient to make the



(a) Free field, $R_b = R_d = 0.3$ m

(b) Free field, $R_b = R_d = 0.5$ m

(c) Reverberant room, $r_p = 0.6$

(d) Reverberant room, $r_p = 0.8$

Fig. 3. (Color online) Best average acoustic contrast level at each frequency.

loudspeaker array emit a sound beam to the bright zone. Therefore, the outperformance of the wave-domain method cannot be demonstrated in this simple situation.

The reverberant cases with $R_b = R_d = 0.3$ m are shown in Figs. 3(c) and 3(d), suggesting the advantage of WDACC at high frequencies. As the reflection coefficient increases from 0.6 to 0.8, the contrast performances of both methods decrease rapidly and the advantage of WDACC becomes more significant. The reduction of spherical harmonics from 5 to 3 does not deteriorate the performance of WDACC in the reverberant environment, but less measurement reduces the high frequency performance of ACC method. This is because the strong reverberation makes the sound field very complicated. A finite number of spherical harmonics can approximately represent the sound field over a large spatial region, and multizone sound can be manipulated by controlling the coefficients captured by a small number of microphones. Since the ACC method only maximizes the contrast between the multiple points, microphones around the control zones are inadequate when reflection and reverberation complicate the sound field inside the target region, and more microphones are essential over large areas at high frequencies.

5. Conclusions

This paper presents a three-dimensional multizone sound field reproduction method with a circular array of first-order loudspeakers, and this method employs the spherical harmonic decomposition-based WDACC approach. The proposed method avoids the Bessel zeros problem in the wave domain for sound reproduction and facilitates three-dimensional sound field control with a single-layer circular array. The free-field and reverberant simulation provide the performance evaluation with respect to the average acoustic contrast over the regions of interest. The results demonstrate that compared with normal ACC method, the proposed WDACC achieves better contrast performance over control regions with well-chosen regularization parameters. Moreover, truncating the spherical harmonics at a lower order such as 5 or 3 instead of $\lceil ekR/2 \rceil$ can accomplish the multizone goal. Thus, much fewer microphones can be used to reproduce multizone sound in the wave domain.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant Nos. 11474306, 11404367, and 11474307.

References and links

[1]J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," The Journal of the Acoustical Society of America **111**(4), 1695–1700 (2002).

[2]Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," IEEE Transactions on audio, speech, and language processing **19**(6), 1711–1720 (2011).

[3]B. Bu, C.-c. Bao, and M.-s. Jia, "Design of a planar first-order loudspeaker array for global active noise control," IEEE/ACM Transactions on Audio, Speech, and Language Processing **26**(11), 2240–2250 (2018).

[4]Z. Han, M. Wu, Q. Zhu, and J. Yang, "Two-dimensional multizone sound field reproduction using a wave-domain method," The Journal of the Acoustical Society of America **144**(3), EL185–EL190 (2018).

[5]E. G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*, Chap. 6, 186,200 (Academic press).

[6]M. A. Wieczorek and M. Meschede, "Shtools: Tools for working with spherical harmonics," Geochemistry, Geophysics, Geosystems **19**(8), 2574–2592 (2018).

[7]B. Rafaely, *Fundamentals of spherical array processing*, Vol. 8, Chap. 2, 45–47 (Springer).

[8]M. Poletti, F. Fazi, and P. Nelson, "Sound-field reproduction systems using fixed-directivity loudspeakers," The Journal of the Acoustical Society of America **127**(6), 3590–3601 (2010).

[9]I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Chap. 8, 964 (Academic press).

[10]R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," IEEE Transactions on Signal processing **55**(6), 2542–2556 (2007).

[11]J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," The Journal of the Acoustical Society of America **65**(4), 943–950 (1979).

[12]J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE (2002), Vol. 2, pp. 1781–1784.