2.5D multizone reproduction using weighted mode matching: Performance analysis and experimental validation

Junqing Zhang, Wen Zhang, Thushara D. Abhayapala, and Lijun Zhang

Citation: The Journal of the Acoustical Society of America **147**, 1404 (2020); doi: 10.1121/10.0000797
View online: https://doi.org/10.1121/10.0000797
View Table of Contents: https://asa.scitation.org/toc/jas/147/3
Published by the Acoustical Society of America

JASA ARTICLE

2.5D multizone reproduction using weighted mode matching: Performance analysis and experimental validation

Junqing Zhang,[1] Wen Zhang,[1,a)] Thushara D. Abhayapala,[2] and Lijun Zhang[1]

[1]*Center of Intelligent Acoustics and Immersive Communications, School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an Shaanxi 710072, China*

[2]*Research School of Electrical, Energy and Materials Engineering, College of Engineering and Computer Science, The Australian National University, Canberra ACT 2601, Australia*

ABSTRACT:

Mode-matching based multizone reproduction has been mainly focused on a purely two-dimensional (2D) theory, where infinite-long 2D secondary sources are assumed for 2D multizone reproduction. Its extension to the three-dimensional (3D) case requires more secondary sources and a higher computational complexity. This work investigates a more practical setup to use 3D sound sources as secondary sources for multizone reproduction in a 2D horizontal plane, i.e., 2.5D multizone reproduction. A weighted mode-matching approach is proposed to solve the dimensionality mismatch between the 2D desired sound field and 3D reproduced sound field. The weighting is based on an integral of Bessel-spherical harmonic modes over the entire control region. A detailed analysis of the weighting function is provided to show that the proposed method controls all the reproduction modes present on the 2D plane to minimize the reproduction error. The method is validated in both simulation-based and hardware-based experiments. The results demonstrate that in comparison with the conventional sectorial mode-matching method, the proposed approach can achieve more accurate reproduction over a wide frequency range and a large control region. © *2020 Acoustical Society of America*.

https://doi.org/10.1121/10.0000797

(Received 13 August 2019; revised 5 February 2020; accepted 6 February 2020; published online 2 March 2020)

[Editor: Julien de Rosny] Pages: 1404–1417

I. INTRODUCTION

Today, the development of multizone sound field reproduction has gained considerable research interest. The aim of multizone reproduction is to provide a variety of listening experiences to different listeners in a shared space, such as listening to music, watching movies, or even generating quiet zones simultaneously in designated areas using a single loudspeaker array.[1,2] This technology provides significant flexibility because there is no need for listeners to wear headphones or be physically isolated, leading to a wide range of audio applications.[3–5]

Up to now, many approaches have been developed for multizone reproduction, including frequency-domain,[6–8] time-domain,[9–11] and modal-domain[12–16] methods. The first two are based on multi-point control that achieves optimal reproduction at specific points. The modal-domain method, on the other hand, decomposes the sound field on a spatial orthogonal basis and then uses the mode-matching approach to control reproduction over a region.

The mode-matching approach, however, has been mainly considered in the two-dimensional (2D) case, i.e., both the virtual source generating the desired sound field and the secondary source used for reproduction are modelled as 2D near-field or far-field sources. However, the theory can be easily extended to the three-dimensional (3D) case. Recently, a weighted mode-matching (WMM) based approach has been presented for 3D sound field reproduction using a 3D loudspeaker array, that is both the desired field and reproduced field are 3D in nature.[17] The optimal weights are derived on the expansion coefficients according to the norm to be minimized. Given the same reproduction setup, the 3D approach that controls sound within a sphere requires more secondary sources and a higher computational complexity.

We investigate the problem of controlling 2D multizone sound fields using 3D point sources in this work, which is a more practical setup because, on the one hand, the same number of loudspeakers as in the 2D reproduction is required, and on the other hand, loudspeakers are more accurately modelled as 3D point sources.[18] However, there is an intrinsic dimensionality mismatch between the 2D desired sound fields and 3D secondary sources, also known as 2.5D reproduction.[19,20] In order to solve the dimensionality mismatch problem, the sectorial mode-matching (SMM) approach[21] was proposed, where the desired and generated sound fields are decomposed through spatial harmonics and the matching at the center of the setup is applied. Analytical solutions were derived for a circular array of secondary sources.[22–24] This approach, however, has mainly been verified for the case of a single sound zone, which is also located at the reproduction center.

In this paper, we further explore the previously proposed WMM approach for 2.5D multizone reproduction

[a)]Electronic mail: wen.zhang@nwpu.edu.cn, ORCID: 0000-0002-0752-6123.

JASA
https://doi.org/10.1121/10.0000797

based on the concept outlined in previous work.[15,25] Instead of referencing the synthesized sound field to the desired one at a particular point or radius, is that the matching is over the entire control region. This is based on weighting the Bessel-spherical harmonic modes. The key contributions of this work are: (i) a detailed analysis of the weight function is provided, and (ii) simulation-based and hardware-based experiments are performed to validate the proposed method.

The structure of this paper is as follows. In Sec. II, we review the theory of 2D modal-domain multizone reproduction and adopt the semidefinite relaxation (SDR) approach to obtain the 2D global sound field coefficients. In Sec. III, the WMM approach is proposed for 2.5D reproduction when the secondary sources for reproduction are assumed as 3D point sources. We provide a detailed analysis of the weighting to solve the dimensionality mismatch problem. The simulation-based and hardware-based experiments are given in Secs. IV and V, respectively. We present the comparison between the proposed WMM and the conventional SMM method, where the experimental results show that the proposed method can achieve better reproduction performance over a wide frequency range and a large control region. Finally, the summation of this work is concluded in Sec. VI.

II. 2D MULTIZONE REPRODUCTION

As a starting point, we review the 2D multizone reproduction formulation as in Ref. 25. As shown in Fig. 1, we assume that each sound zone q has a radius r_q and its centre is denoted by o_q with respect to the global origin o. Any observation point within a sound zone is represented by $\mathbf{x}_q$ with respect to o_q, or $\mathbf{x} = \mathbf{x}_q + \mathbf{o}_q$ with respect to o. All sound zones are within a general region of interest with radius $r \le r_0$.

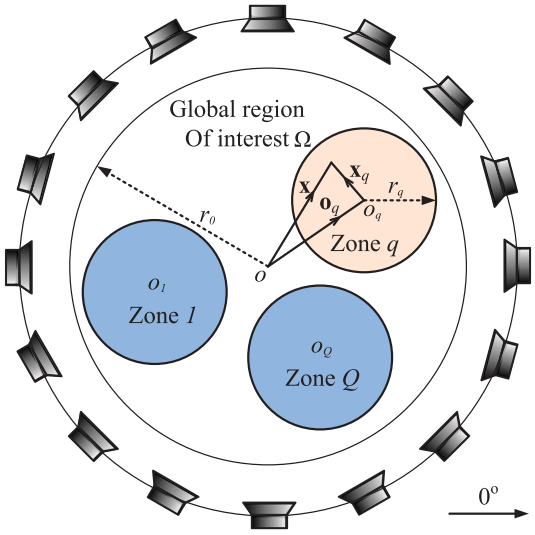

FIG. 1. (Color online) Geometry of multizone sound field control.

A. Sound field model

A 2D source-free incoming field at an arbitrary point $\mathbf{x}_q \equiv (||\mathbf{x}_q||, \phi_{\mathbf{x}_q})$ can be approximately represented in terms of a harmonic decomposition of the form[26]

$$P^{(q)}(\mathbf{x}_q, k) \approx \sum_{m'=-N_q}^{N_q} \alpha_{m'}^{(q)}(k) J_{m'}(k||\mathbf{x}_q||) e^{im'\phi_{\mathbf{x}_q}}, \tag{1}$$

where $k = 2\pi f/c$ is the wave number with f being the frequency and c the speed of sound, $J_{m'}(k||\mathbf{x}_q||)$ is the cylindrical Bessel function of order m', and $\alpha_{m'}^{(q)}(k)$ is the corresponding m' th order sound field coefficient to describe a spatial sound field with respect to o_q. Given the radius of the local sound zone r_q, the wave number k, the truncation order is $N_q = \lceil ekr_q/2 \rceil$,[27] where $\lceil \cdot \rceil$ denotes the ceiling function and e is the Euler's number.

Similar to Eq. (1), the global sound field at the point $\mathbf{x} \equiv (||\mathbf{x}||, \phi_{\mathbf{x}})$ can be written as

$$P(\mathbf{x}, k) \approx \sum_{m=-N_0}^{N_0} \beta_m(k) J_m(k||\mathbf{x}||) e^{im\phi_{\mathbf{x}}}, \tag{2}$$

where the truncation order $N_0 = \lceil ekr_0/2 \rceil$ and r_0 is the radius of the general region of interest.

The local sound field coefficients $\alpha_{m'}^{(q)}(k)$ can be related to the global sound field coefficients $\beta_m(k)$ using the 2D Bessel function addition theorem,[28] that is

$$J_m(k||\mathbf{x}||)e^{im\phi_{\mathbf{x}}} = \sum_{m'=-N_q}^{N_q} J_{m-m'}(k||\mathbf{o}_q||)e^{i(m-m')\phi_{\mathbf{o}_q}} J_{m'}(k||\mathbf{x}_q||)e^{im'\phi_{\mathbf{x}_q}}, \tag{3}$$

given $\mathbf{x} = \mathbf{x}_q + \mathbf{o}_q$ and $\mathbf{o}_q \equiv \{||\mathbf{o}_q||, \phi_q\}$ in the global system.

From Eqs. (1)–(3), we can represent the relationship between global and local region in a compact form as

$$\boldsymbol{\alpha}_q = \mathbf{T}_q \boldsymbol{\beta}, \tag{4}$$

where $\boldsymbol{\alpha}_q = [\alpha_{-N_q}^{(q)}(k), \ldots, \alpha_{N_q}^{(q)}(k)]^T$ and $\boldsymbol{\beta} = [\beta_{-N_0}(k), \ldots, \beta_{N_0}(k)]^T$ are column vectors of length $(2N_q + 1)$ and $(2N_0 + 1)$, respectively. $\mathbf{T}_q$ is the $(2N_q + 1) \times (2N_0 + 1)$ translation matrix between the global region and the local region, that is $[\mathbf{T}_q]_{m'+N_q+1, m+N_0+1} = J_{m-m'}(k||\mathbf{o}_q||)e^{i(m-m')\phi_{\mathbf{o}_q}}$.

B. Problem formulation and solution using SDR

Modal-domain control of the multizone reproduction problem can be formulated by finding the global sound field coefficients $\boldsymbol{\beta}$ to reproduce a desired sound field in the bright zone Ω_b characterised by its local coefficients $\boldsymbol{\alpha}_b$ with constraints on the sound energy in the dark zone Ω_d and the total energy of the global sound field,[29] that is

$$\min_{\boldsymbol{\beta}} \ ||\mathbf{T}_b\boldsymbol{\beta} - \boldsymbol{\alpha}_b||^2 \tag{5a}$$

$$\text{subject to} \quad ||\mathbf{T}_\text{d}\boldsymbol{\beta}||^2 \le e_\text{d}, \tag{5b}$$

$$||\boldsymbol{\beta}||^2 \le e_\text{g}, \tag{5c}$$

where $\mathbf{T}_\text{b}$ and $\mathbf{T}_\text{d}$ are the translation matrices of the local bright zone and dark zone from the global system, respectively. In terms of the two energy terms of Eqs. (5b) and (5c), e_d is determined by a threshold on the sound level within the dark zone, and e_g is the energy of the entire global sound field. For personal audio, perceptual requirements can also be taken into account for setting the values of e_d and e_g, such as the noticeable level of reproduction error within the bright zone and the acceptable level of audio interference reduction within the dark zone.[29]

The above formula can be classified as a complex-valued quadratically constrained quadratic program (QCQP) problem. As in Refs. 30 and 31, Eq. (5) can be converted to a real-valued form as follows:

$$\mathbf{y} = \begin{bmatrix} \text{Re}\{\boldsymbol{\alpha}_\text{b}\} \\ \text{Im}\{\boldsymbol{\alpha}_\text{b}\} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \text{Re}\{\boldsymbol{\beta}\} \\ \text{Im}\{\boldsymbol{\beta}\} \\ t \end{bmatrix},$$

$$\mathbf{S}_\text{b} = \begin{bmatrix} \text{Re}\{\mathbf{T}_\text{b}\} & -\text{Im}\{\mathbf{T}_\text{b}\} \\ \text{Im}\{\mathbf{T}_\text{b}\} & \text{Re}\{\mathbf{T}_\text{b}\} \end{bmatrix},$$

$$\mathbf{S}_\text{d} = \begin{bmatrix} \text{Re}\{\mathbf{T}_\text{d}\} & -\text{Im}\{\mathbf{T}_\text{d}\} \\ \text{Im}\{\mathbf{T}_\text{d}\} & \text{Re}\{\mathbf{T}_\text{d}\} \end{bmatrix},$$

$$\mathbf{A}_\text{b} = \begin{bmatrix} \mathbf{S}_\text{b}^T\mathbf{S}_\text{b} & -\mathbf{S}_\text{b}^T\mathbf{y} \\ -\mathbf{y}^T\mathbf{S}_\text{b} & \mathbf{y}^T\mathbf{y} \end{bmatrix}, \quad \mathbf{A}_\text{d} = \begin{bmatrix} \mathbf{S}_\text{d}^T\mathbf{S}_\text{d} & 0 \\ 0 & 0 \end{bmatrix},$$

where t, which satisfies $t^2 = 1$, is an extra variable to homogenize the Eq. (5).[31]

Then, the problem can be reposed as homogeneous QCQP

$$\min_{\mathbf{x}} \quad \mathbf{x}^T\mathbf{A}_\text{b}\mathbf{x}, \tag{6a}$$

$$\text{subject to} \quad \mathbf{x}^T\mathbf{A}_\text{d}\mathbf{x} \le e_\text{d}, \tag{6b}$$

$$\mathbf{x}^T\mathbf{x} \le e_\text{g}, \tag{6c}$$

$$t^2 = 1. \tag{6d}$$

Instead of solving the Lagrangian solution to the cost functions of Eq. (5),[25] we choose to use SDR, a computationally efficient technique for solving quadratic optimization problems,[31] to find the approximated solution of Eq. (6).

By defining a variable $\mathbf{X} = \mathbf{x}\mathbf{x}^T$, we can obtain the SDR of Eq. (6) as

$$\min_{\mathbf{X}} \quad Tr(\mathbf{A}_\text{b}\mathbf{X}), \tag{7a}$$

$$\text{subject to} \quad Tr(\mathbf{A}_\text{d}\mathbf{X}) \le e_\text{d}, \tag{7b}$$

$$Tr(\mathbf{X}) \le e_\text{g}, \tag{7c}$$

$$\mathbf{X} \succcurlyeq 0, \tag{7d}$$

$$t^2 = 1, \tag{7e}$$

where the notation $\succcurlyeq$ means the matrix $\mathbf{X}$ is positive semidefinite. The detailed derivation is shown in Appendix A. By solving this SDR problem, we can obtain the global coefficients $\boldsymbol{\beta}$. Note that the SDR has polynomial running time, and it can be handled conveniently using readily available software packages, such as CVX. We refer readers to Ref. 31 for a detailed derivation and applications of the theory.

III. 2.5D REPRODUCTION USING WEIGHTED MODE MATCHING

The problem of 2.5D reproduction is to solve the dimensionality mismatch between the 2D desired sound fields and 3D secondary sources. In the existing methods, the matching point is simply at the center of the setup, i.e., the global origin at $r = 0$,[21] i.e., the SMM approach. However, for multi-zone reproduction, local sound zones are normally away from the center. We propose a matching over the entire global region, i.e., the reference distance $r \in [0, r_0]$, using the weighting function as shown in the following.

A. Algorithm 1: Weighted mode matching

The WMM is based on the following cost function to minimize the reproduction error over the entire control region Ω[25]

$$\mathcal{J}(\mathbf{d}, k) = \frac{1}{2\pi}\int_\Omega |\hat{P}(\mathbf{x}, k) - P_\text{des}(\mathbf{x}, k)|^2 d\sigma(\mathbf{x}), \tag{8}$$

where $d\sigma(\mathbf{x}) = rdrd\phi_\mathbf{x}$ is the differential area element at $\mathbf{x}$, and $r = ||\mathbf{x}||$.

Referring to Eq. (2), the global desired sound field can be represented as

$$P_\text{des}(\mathbf{x}, k) \approx \sum_{m=-N_0}^{N_0} \underbrace{\beta_m(k)J_m(k||\mathbf{x}||)}_{b_m(||\mathbf{x}||,k)} e^{im\phi_\mathbf{x}}. \tag{9}$$

The global sound field generated by an array of L loudspeakers can be written as

$$\hat{P}(\mathbf{x}, k) = \sum_{\ell=1}^{L} d_\ell(k)G_\ell(\mathbf{x}, k), \tag{10}$$

where $G_\ell(\mathbf{x}, k)$ represents the acoustic transfer function (ATF) between the ℓ th loudspeaker and the observation point $\mathbf{x}$ in the global system, and $[\mathbf{d}]_\ell = d_\ell$ is the ℓ th loudspeaker driving signal.

The ATF for a 3D source is parameterised in the modal domain as

$$G_\ell(\mathbf{x}, k) \approx \sum_{m=-N_0}^{N_0} \sum_{n=|m|}^{N_0} \gamma_n^m(\ell, k) j_n(k||\mathbf{x}||) Y_n^m(\theta_\mathbf{x}, \phi_\mathbf{x}), \tag{11}$$

where $Y_n^m(\theta_\mathbf{x}, \phi_\mathbf{x})$ is the spherical harmonic function, j_n is the nth order spherical Bessel function of the first kind, and

JASA
https://doi.org/10.1121/10.0000797

$\gamma_n^m(\ell,k)$ are ATF coefficients for the ℓ th speaker. Note that each loudspeaker is a 3D point source, there are $(N_0+1)^2$ coefficients to describe its ATF with respect to the global control region. The ATF coefficients are assumed to be prior knowledge, obtained from theoretical solutions or pre-calibration.[32] In anechoic condition $\gamma_n^m(\ell,k) = -ikh_n^{(2)}(kr_\ell)\overline{Y_n^m(\theta_\ell,\phi_\ell)}$, where $\overline{(\cdot)}$ represents the complex conjugate, $h_n^{(2)}$ is spherical Hankel function of second kind, and $\mathbf{x}_\ell \equiv (r_\ell,\theta_\ell,\phi_\ell)$ is the position of the sources.

Then, the generated 3D sound field at the height of the human ear (elevational angle $\theta_l = \pi/2$) is

$$\hat{P}(\mathbf{x},k) \approx \sum_{m=-N_0}^{N_0}\sum_{\ell=1}^{L} d_\ell(k) \underbrace{\sum_{n=|m|}^{N_0} \gamma_n^m(\ell,k) j_n(k||\mathbf{x}||)Q_n^m}_{h_m(\ell,||\mathbf{x}||,k)} e^{im\phi_\mathbf{x}}, \tag{12}$$

where $Y_n^m(\pi/2,\phi_\mathbf{x}) = Q_n^m e^{im\phi_\mathbf{x}}$ with $Q_n^m = (-1)^m A_n^m P_n^{|m|}(0)$ and $A_n^m = \sqrt{[(2n+1)/(4\pi)][(n-|m|)!/(n+|m|)!]}$.

Substituting Eqs. (9) and (12) into Eq. (8) leads to

$$\mathcal{J}(\mathbf{d},k) = \sum_{m=-N_0}^{N_0}\int_0^{r_0}\left|\sum_{\ell=1}^{L} d_\ell(k)h_m(\ell,r,k) - b_m(r,k)\right|^2 rdr. \tag{13}$$

We then write Eq. (13) in matrix form

$$\mathcal{J}(\mathbf{d},k) = \mathbf{d}^H\mathbf{H}\mathbf{d} - \mathbf{B}^H\mathbf{d} - \mathbf{d}^H\mathbf{B} + C, \tag{14}$$

with

$$C = \int_0^{r_0}\mathbf{b}(r,k)^H\mathbf{b}(r,k)rdr, \tag{15}$$

where the desired field vector $\mathbf{b}(r,k) = [b_{-N_0}(r,k),\ldots,b_{N_0}(r,k)]^T$.

$$[\mathbf{H}]_{\ell1,\ell2} = \sum_{m=-N_0}^{N_0}\sum_{n=|m|}^{N_0}\sum_{n'=|m|}^{N_0}\omega_{n,n'}^m\overline{\gamma_n^m(\ell_1,k)}\gamma_{n'}^m(\ell_2,k).$$

For notation clarity, we define both n and n' as the degree index

$$[\mathbf{B}]_{\ell1} = \sum_{m=-N_0}^{N_0}\sum_{n=|m|}^{N_0}\chi_n^m\overline{\gamma_n^m(\ell_1,k)}\beta_m(k),$$

and

$$\omega_{n,n'}^m \triangleq Q_n^m Q_{n'}^m\int_0^{r_0} j_n(kr)j_{n'}(kr)rdr,$$
$$\chi_n^m \triangleq Q_n^m\int_0^{r_0} j_n(kr)J_m(kr)rdr.$$

These two integrals are continuous and are numerically calculated in this work using a trapezoidal method with 10^{-4} spacing.

We further write

$$\mathbf{H} = \mathbf{\Gamma}^H\mathbf{W}\mathbf{\Gamma}, \tag{16}$$

where $\mathbf{\Gamma}$ is the ATF coefficient matrix defined as $[\mathbf{\Gamma}]_{n^2+n+m+1,\ell} = \gamma_n^m(\ell,k)$ and $\mathbf{W}$ is a $(N+1)^2$-square weighting matrix defined as

$$[\mathbf{W}]_{n^2+n+m+1,n'^2+n'+m'+1} = \delta_{m-m'}\omega_{n,n'}^m. \tag{17}$$

The superscript $[\cdot]^H$ denotes the conjugate transpose operator. Note that here we define both m and m' as the order index.

Similarly, it can be defined that $\mathbf{B} = \mathbf{\Gamma}^H\mathbf{X}\boldsymbol{\beta}$, where

$$[\mathbf{X}]_{n^2+n+m+1,m'+N_0+1} = \delta_{m-m'}\chi_n^m,$$

and $[\boldsymbol{\beta}]_{m+N_0+1} = \beta_m(k)$.

By solving the cost function in Eq. (14), we can obtain the loudspeaker driving signals

$$\hat{\mathbf{d}} = \mathbf{H}^{-1}\mathbf{B} = (\mathbf{\Gamma}^H\mathbf{W}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^H\mathbf{X}\boldsymbol{\beta}. \tag{18}$$

B. Algorithm 2: Sectorial weighted mode matching

SMM only matches the sectorial modes at $n = |m|$ and the matching point is at the global origin $r = 0$.[21] Here, we remove the requirement of the matching at the center and apply the weighting approach to the SMM. That is, for the sectorial mode approximation,

$$h_m^{\text{sect}}(\ell,||\mathbf{x}||,k) = \gamma_{|m|}^m(\ell,k)j_{|m|}(k||\mathbf{x}||)Q_{|m|}^m. \tag{19}$$

Then we have

$$\mathbf{H}_{\text{sect}} = \mathbf{\Gamma}_{\text{sect}}^H\mathbf{W}_{\text{sect}}\mathbf{\Gamma}_{\text{sect}}, \tag{20}$$

where $\mathbf{\Gamma}_{\text{sect}}$ is a matrix of size $(2N_0+1)\times L$, i.e., $[\mathbf{\Gamma}_{\text{sect}}]_{m+N_0+1,\ell} = \gamma_{|m|}^m(\ell,k)$, and the diagonal weighting matrix

$$[\mathbf{W}_{\text{sect}}]_{m+N_0+1,m+N_0+1} = |Q_{|m|}^m|^2\int_0^{r_0}|j_{|m|}(kr)|^2rdr,$$

is of size $(2N_0+1)\times(2N_0+1)$.

Given the global sound field coefficient $\boldsymbol{\beta}$ and

$$[\mathbf{X}_{\text{sect}}]_{m+N_0+1,m+N_0+1} = Q_{|m|}^m\int_0^{r_0} j_{|m|}(kr)J_m(kr)rdr,$$

the solution for the sectorial weighted mode matching (SWMM) is

$$\hat{\mathbf{d}}_{\text{sect}} = \mathbf{H}_{\text{sect}}^{-1}\mathbf{B}_{\text{sect}} = (\mathbf{\Gamma}_{\text{sect}}^H\mathbf{W}_{\text{sect}}\mathbf{\Gamma}_{\text{sect}})^{-1}\mathbf{\Gamma}_{\text{sect}}^H\mathbf{X}_{\text{sect}}\boldsymbol{\beta}. \tag{21}$$

The weighting terms are introduced to solve the dimensionality mismatch problem in 2.5D reproduction. The traditional method only matches the 2D desired field and 3D

reproduced field at the origin, i.e., achieving the optimal control at this point. The weighting terms we proposed have an integral of Bessel-spherical harmonic modes over the entire global region; based on this, we can control all the reproduction modes present on the 2D plane and thus can achieve optimal control over the entire region.

C. Performance analysis of weighting function

In this subsection, we analyse the weighting matrix **W** and the Bessel integral term within **W**. In addition, we compare the three modal-domain 2.5D reproduction methods, i.e., the conventional SMM,[21] the proposed WMM, and SWMM, and have the following remarks.

- The weighting matrix **W** as shown in Fig. 2 is a symmetric matrix. It is noted that at the horizontal plane when $n+|m|$ is an odd integer, the associated Legendre functions $P_n^{|m|}(0)$ in Q_n^m are equal to zero.[33,34] Therefore, there are many zeros in **W**, and the non-zero elements are mainly concentrated on principal diagonal lines. In Fig. 2 different orders are separated by the white line. The low order modes have bigger values comparing to the higher order modes, showing that the weighting has a greater control effect on low order modes that becomes effective at smaller r (or more prevalent over the entire control region).
- The weighting matrix **W** in Eq. (17) involves computing the integral of Bessel-spherical harmonic modes. We define the Bessel integral term $E(kr)$ to investigate the influence of the Bessel function term in **W**, i.e.,

$$E(kr) = \int_0^{r_0} j_n(kr) j_{n'}(kr) r dr. \tag{22}$$

Figure 3 shows the results of $E(kr)$ of different orders along with kr. Note that the zeroth order mode, which has the biggest value, always exists, while higher order modes gradually appear as kr increases. This also demonstrates

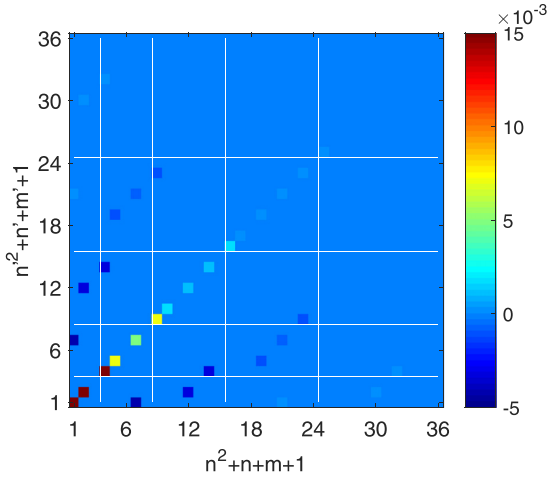

FIG. 2. (Color online) A plot of the weighting matrix **W** with index $n^2+n+m+1$, $n=0,\ldots,N_0$, $m=-n,\ldots,n$, where $\lceil ekr/2 \rceil = 5$.

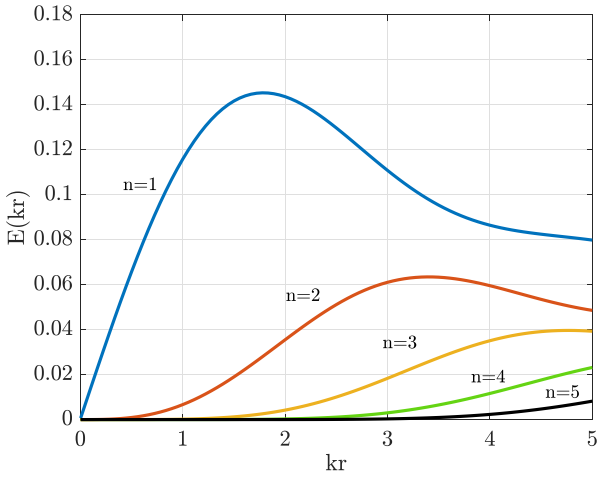

FIG. 3. (Color online) A plot of the Bessel integral term $E(kr)$ for $n=1,\ldots,5$ as a function of kr.

that the lower order modes are more prevalent over the entire region while higher order modes become effective for larger r.

- The biggest difference in the proposed weighting methods and conventional SMM method is the number of effective control modes. As shown in Fig. 4, the black dots, black circles, and white dotted circles correspond to the sectorial, elevational, and zero modes, respectively,[35] while the n and m in x and y axes represent the different degree and order. The SMM and SWMM approaches only match the sectorial modes at $n=|m|$. Thus, the number of the control modes is $(2N_0+1)$. For the WMM method truncated to N_0, out of the $(N_0+1)^2$ modes, the number of non-zero terms is $(N_0/2+1)(N_0/4+1)$ when N_0 is even and $[(N_0-1)/2+1][(N_0-1)/4+1]$ when N_0 is odd. Comparing to the SMM and SWMM, the WMM guarantees that all modes present on the horizontal plane are controlled to minimize the reproduction error within the entire control region.

IV. SIMULATION

In this section, we verify the effectiveness of the proposed reproduction methods through simulation-based experiments. We make the comparison between the proposed WMM/SWMM methods and the existing SMM method.[21]

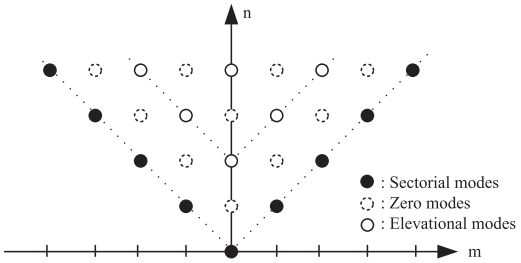

FIG. 4. Sectorial and elevational modes in the x–y plane.

JASA
https://doi.org/10.1121/10.0000797

A. Simulation setup

We simulated two-zone reproduction examples under free field and reverberant environments. A rectangular room of size $5\,\mathrm{m} \times 7$ m is simulated using the image source method[36] with the image order up to 5, the wall reflection coefficients up to 0.7, and the floor and ceiling being perfectly-absorbing. The radius of the global control region is $r_0 = 0.8$ m. The virtual source, which is located in the far field and incident from $\phi_V = \pi/3$, produces a monochromatic plane wave of frequency 1000 Hz. The bright zone and dark zone are located at $\mathbf{o}_\mathrm{b} = (0.5, 0)$ and $\mathbf{o}_\mathrm{d} = (-0.5, 0)$ with respect to the global origin, respectively. Each sound zone has a radius of 0.15 m. A 41-element circular loudspeakers array of radius 1 m is used for reproduction. The proposed SDR method mentioned in Sec. II B is used to solve the desired global coefficients $\boldsymbol{\beta}$. By referring to Eqs. (5b) and (5c), we set the energy constraints in the dark zone and global region as $e_\mathrm{d} = -10$ dB and $e_\mathrm{g} = 20$ dB for perceptual requirements in personal audio.[29]

Three evaluation indicators are given as follows:

- The acoustic contrast (AC) $\kappa(k)$ between the bright zone and dark zone

$$\kappa(k) = 10\log_{10} \frac{\frac{1}{\mathrm{V_b}} \int_{\Omega_\mathrm{b}} |\hat{P}(\mathbf{x}, k)|^2 d\sigma(\mathbf{x})}{\frac{1}{\mathrm{V_d}} \int_{\Omega_\mathrm{d}} |\hat{P}(\mathbf{x}, k)|^2 d\sigma(\mathbf{x})}. \tag{23}$$

- The bright zone reproduction error $\varepsilon(k)$

$$\varepsilon(k) = 10\log_{10} \frac{\frac{1}{\mathrm{V_b}} \int_{\Omega_\mathrm{b}} |\hat{P}(\mathbf{x}, k) - P_\mathrm{des}(\mathbf{x}, k)|^2 d\sigma(\mathbf{x})}{\frac{1}{\mathrm{V_b}} \int_{\Omega_\mathrm{b}} |P_\mathrm{des}(\mathbf{x}, k)|^2 d\sigma(\mathbf{x})}. \tag{24}$$

- The array effort or the loudspeaker weight energy to quantify the total output for generating the desired sound field

$$\iota(k) = 10\log_{10} ||\hat{\mathbf{d}}(k)||^2, \tag{25}$$

where $\hat{P}(\mathbf{x}, k)$ and $P_\mathrm{des}(\mathbf{x}, k)$ represent the reproduced sound field and the desired sound field at a point $\mathbf{x}$ within the bright zone Ω_b or the dark zone Ω_d. $\mathrm{V_b}$ and $\mathrm{V_d}$ denote the area of the bright zone and the dark zone, respectively.

B. Two-zone examples

The real part of the reproduced results are plotted in Fig. 5. The display is limited to the maximum value of the reconstructed field within the bright zone, i.e., the acoustic pressures greater than 1 are white and those less than -1 are black. Reproduction results in free field using SMM, SWMM, and WMM are shown in Figs. 5(a), 5(b), and 5(c). The AC of these three methods is 25.59, 21.65, and 25.83 dB, respectively. The bright zone reproduction errors are -9.21, -13.00, and -27.40 dB, respectively. The WMM method demonstrates the highest AC and the lowest

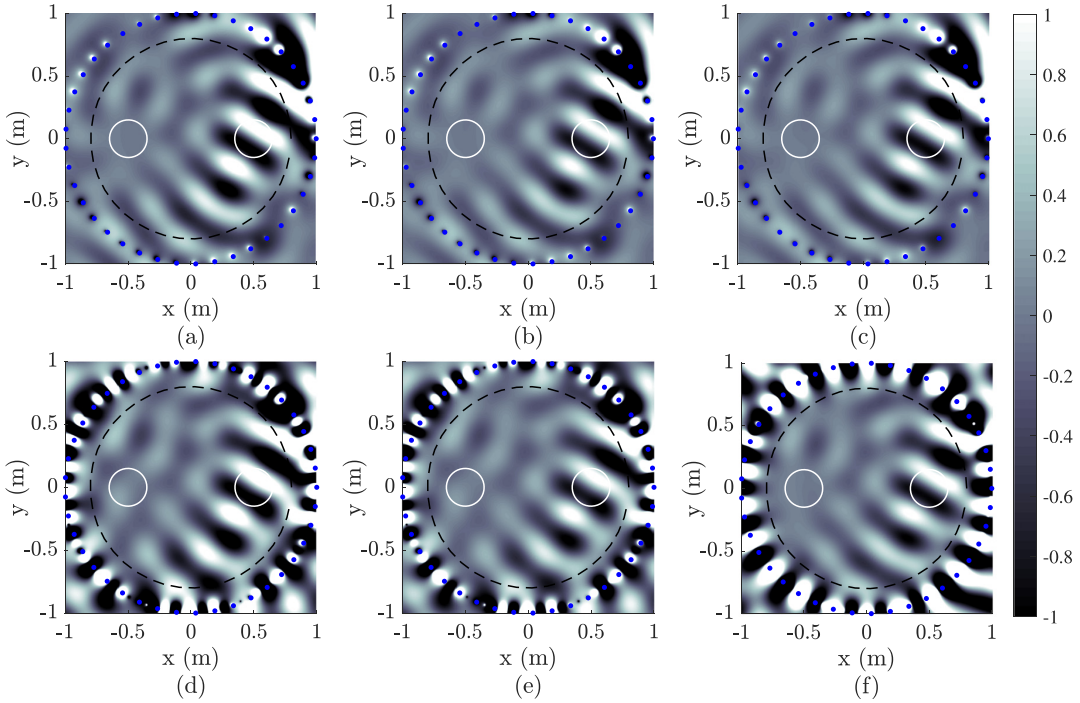

FIG. 5. (Color online) Examples of a two-zone reproduction where a plane wave of frequency 1000 Hz from $\pi/3$ is reproduced in the bright zone. Blue · marks denote the locations of loudspeakers. The area in the black dashed line and white solid line correspond to the global region, bright zone, and dark zone, respectively. First row: (a) SMM, (b) SWMM, (c) WMM in free field. Second row: (d) SMM, (e) SWMM, (f) WMM in reverberation room.

bright zone reproduction error. The loudspeaker weight energy of these three methods is 15.08, 6.94, and 13.28 dB, respectively.

The second row in Fig. 5 shows the reproduction results of these three methods in the reverberant environment. The AC for SMM, SWMM, and WMM is 17.74, 17.07, and 31.76 dB, and the corresponding bright zone reproduction errors are −9.40, −12.18, and −26.67 dB, respectively. The WMM method also demonstrates the highest AC and the lowest bright zone reproduction error. However, the loudspeaker weight energy required in the WMM method is about 15 dB higher than that required in the SWMM and SMM methods. The reason for this result is that the WMM approach controls a larger number of modes as stated in Sec. III C, thus requiring more loudspeaker energy.

C. Performance comparison of SMM, SWMM, and WMM

In this subsection, we show the evaluation results for comparison of the three reproduction methods. The reproduction setup is the same as in the example of Fig. 5. We especially investigate the performance for a variable of operating frequencies, sound zone locations, and the virtual source directions.

First, we evaluate the system performance over a broadband frequency range of [0.1,2] kHz in free field and a reverberation room. The bright zone and dark zone are about 0.5 m away from the global center and the desired sound field is due to a plane wave from 50°. The blue, black, and red lines correspond to the SMM, SWMM, and WMM methods, respectively. The results in Fig. 6(a) show that the SMM and WMM have roughly the same AC performance, and the SWMM has the worst AC performance. In Fig. 6(d), the WMM method demonstrates the lowest reproduction error. This is due to the fact that WMM applies the model weighting for all the modes present on the horizontal plane, thus the optimal control is achieved over the entire global region.

Next, we evaluate the system performance under different values of the distance between the bright/dark zone center and the global origin. The source operates at the frequency of 1000 Hz and the desired sound field is due to a plane wave from 50°. The results are shown in Figs. 6(b) and 6(e). The conventional SMM has the lowest bright zone reproduction error when the distance is less than 0.15 m. This corresponds to the case that the sound zones are close to the global origin, i.e., the matching point at the center of the setup. On the other hand, the proposed WMM has roughly the same AC performance over the entire distance range, but with the lowest bright zone reproduction error when the distance is above 0.24 m. Compared with SMM, the proposed SWMM only shows marginal improvement of bright zone reproduction accuracy when the sound zones are further away from the center but also with the lowest AC.

We then compare the performance of these three methods for the virtual source (or a plane wave) direction over an angle range of [0°, 180°]. The source operates at the frequency of 1000 Hz and the bright zone and dark zone are about 0.5 m away from the global center. As shown in Fig. 6(c), SMM and WMM have roughly the same AC performance, while SWMM has the worst AC performance. In terms of the reproduction error, the proposed WMM method can achieve the lowest reproduction error for the angle range of [0°, 110°].

Figure 7 gives the evaluation results in the reverberation room. The simulation parameters are consistent with those mentioned above. In Figs. 7(a) and 7(d), the AC results of the SMM and SWMM methods are roughly the same, except the results at several frequency points are quite poor. The

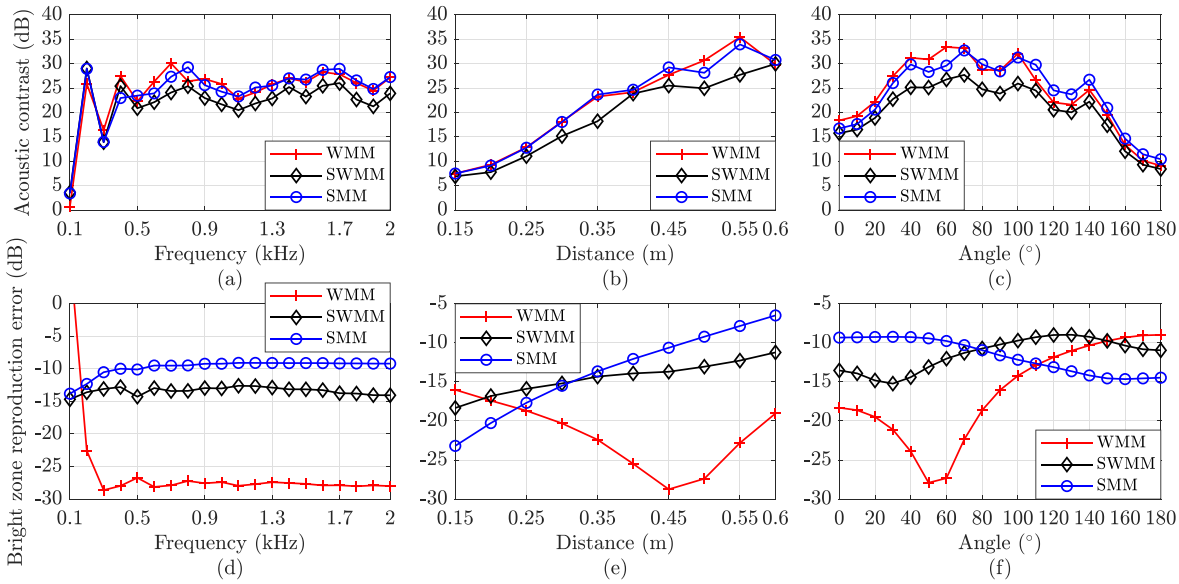

FIG. 6. (Color online) Free field evaluation results. (a), (b), and (c) are the acoustic contrast results varying with frequency, distance, and angle, respectively. (d), (e), and (f) are the bright zone reproduction results varying with frequency, distance and angle, respectively.

https://doi.org/10.1121/10.0000797

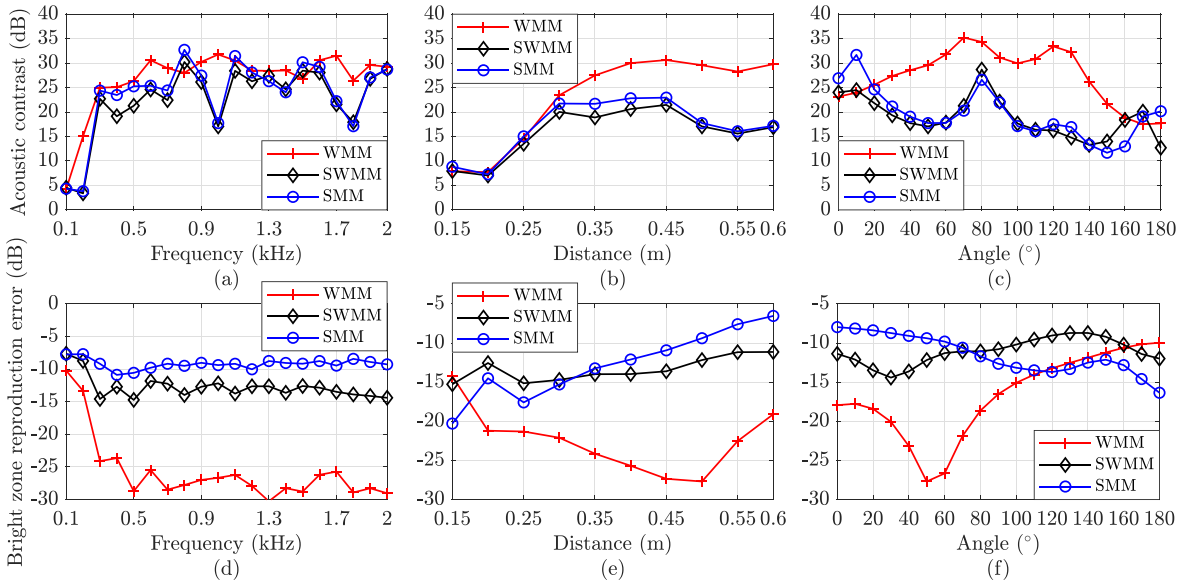

FIG. 7. (Color online) Reverberation room evaluation results. (a), (b), and (c) are the acoustic contrast results varying with frequency, distance, and angle, respectively. (d), (e), and (f) are the bright zone reproduction results varying with frequency, distance, and angle, respectively.

AC performance of the WMM method is slightly better and consistent over the entire frequency band. For sound zones at different distances, WMM also demonstrates significant advantage. The general trend is that the greater the distance, the better WMM performs for both AC and reproduction error as shown in Figs. 7(b) and 7(e). When the virtual source direction changes, the WMM method has better performance over a wide range of angles with the lowest reproduction error reaching -27.88 dB. These results further demonstrate the effectiveness of the proposed WMM method.

V. EXPERIMENTAL VALIDATION

In this section, we provide experimental validation of the proposed weighting methods in comparison with the SMM method. The experimental validation was conducted in the Acoustics and Audio Lab at the Center of Intelligent Acoustics and Immersive Communications (CIAIC), Northwestern Polytechnical University.

A. Experiment setup

We used a circular array of 48 equiangular-placed loudspeakers at the radius 1 m and height 1.65 m to reproduce the desired multizone sound field. The block diagram of equipment electrical connections is shown in Fig. 8. All the devices were controlled under the Dante network. According to the requirement of realization process, the experiment can be divided into ATF measurement and reproduction stages.

1. ATF measurement

The first step is to obtain the ATF modal coefficients $\gamma_n^m(\ell,k)$ as given in Eq. (11). A rigid sphere is preferred to perform this task because it can circumvent the Bessel zeros problem. In our experiment, we used the 32-channel spherical microphone array Eigenmike to measure $\gamma_n^m(\ell,k)$. The Eigenmike is a sphere of aluminium at the radius of 0.042 m with 32 high quality microphones placed on its surface. Elements such as microphones, pre-amplifiers, and A/D converters are all packed inside the sphere, and its operating frequency (i.e., the spatial aliasing frequency) up to 5000 Hz.[37] The systems is capable of recording 32 channel data with 24 bit resolution at a sampling rate of 44.1 or 48 kHz. The clock synchronization between Eigenmike and the Dante network is achieved by connecting the output of the Rednet3 sound card to the EMBI ADAT input of Eigenmike through an ADAT fiber, and by setting Dante as the master network clock.

In the recoding stage, the radius of the global control region is up to 0.6 m corresponding to $(N_0+1)^2 = 256$ modes at 1000 Hz. However, the Eigenmike only can capture an order up to 4, and thus the effective modes number is $(4+1)^2 = 25$. Therefore, we need to combine multiple measurements together to obtain the global ATF coefficients. The minimum requirement of measurement numbers is $Q = \lceil 256/25 \rceil = 11$. We move a single Eigenmike along the boundary of the global control region 11 times at equal angle interval to capture 256-point global sound pressures.[38] Loudspeakers were fed one at a time with a 5-s long exponential sweep of [120, 5000] Hz.[39] Then, the 32-channel captured pressure signals were delivered to the EMBI Firewire interface through a digital CAT-6 cable, and the output of the interfere was transmitted to a Mac Pro for data storage and computing. The ATF coefficient extraction from higher order microphone recordings and translation between the local measurements and global system is shown in Appendix B.

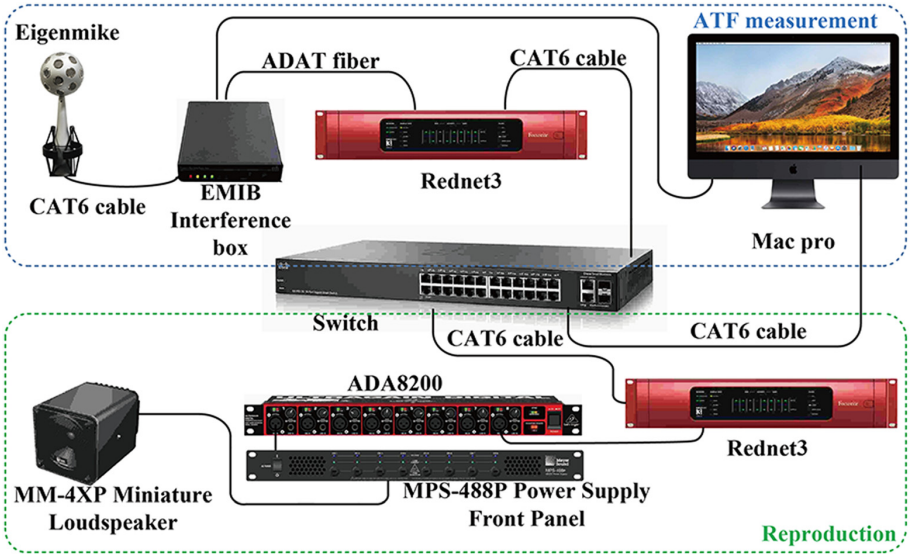

FIG. 8. (Color online) The block diagram of the electrical connections of equipments.

2. Reproduction

Given the ATF coefficients, the next step is to reproduce the desired sound field using the 48-channel loudspeaker array as shown in Fig. 9(a). We use the Meyer Sound MM-4XP miniature loudspeaker for reproduction. This loudspeaker can deliver a maximum peak sound pressure level (SPL) of 113 dB and has a wide operating frequency range from 120 Hz to 18 kHz with very low distortion. The MM-4XP loudspeaker requires a MPS-488 external power supply as shown in Fig. 8. The reproduction system is based on ADAT protocol and 24-bit ADA8200 converters. The sound card we used is a Focusrite Rednet3 at the sampling frequency of 48 kHz. The system is driven by the Dante controller.

As in the simulation section, the global center is set as the origin and we have the bright and dark zone at positive x axis and negative x axis, i.e., 0° and 180°, respectively. The radii of the global control region and local region are 0.6 and 0.15 m, while the latter is comparable to the size of the

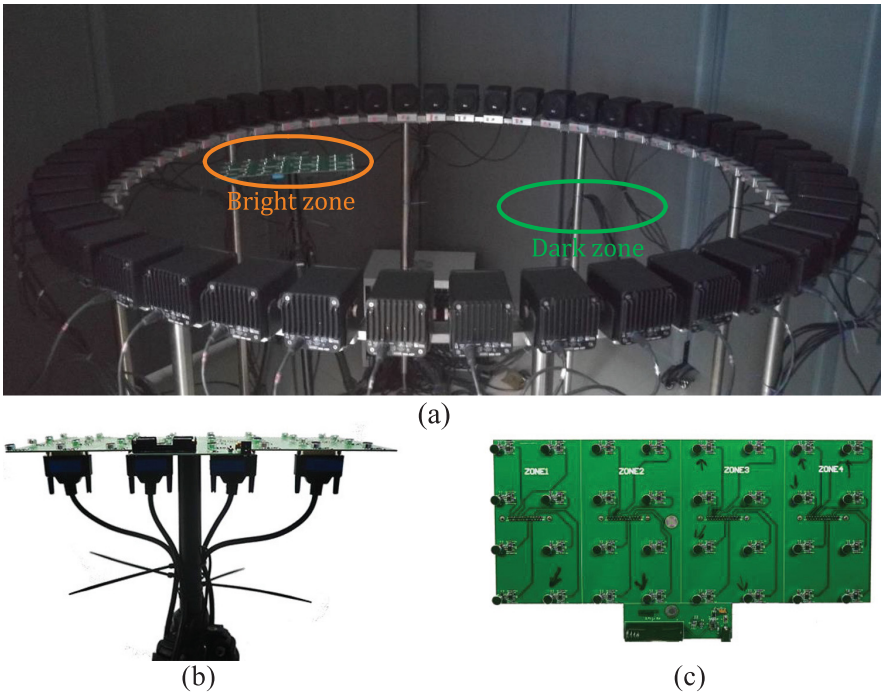

FIG. 9. (Color online) (a) The 48-channel loudspeakers array and 32-channel fixed gain planar Microphone experiment setup enclosed by acoustic baffles. (b) and (c) are the main and top view of the 4×8 elements microphone array.

https://doi.org/10.1121/10.0000797

listener's head. The experimental space within a radius of 1.3 m is enclosed by acoustic baffles (2 m × 0.8 m) which have two surfaces, sound absorption and reverberant, to simulate the semi-anechoic and reverberation acoustic environment. We choose the frequency range of [200, 1000] Hz.

The driving signals were fed to the loudspeakers array to produce an actual sound field, and the sound pressures within the bright and dark region were captured by the 32-channel fixed-gain microphone array placed on a plane as shown in Figs. 9(b) and 9(c). The measurement region of the array is a rectangle of size 0.35 m × 0.15 m, which is able to cover the entire local region. The spacing between the adjacent microphone points is 0.05 m.

B. Results and discussions

The desired field is a plane wave with an amplitude of 0.5 coming from the azimuth angle $\phi = 30^\circ$. In this experiment, a circular array of 48 loudspeakers was used to synthesize this sound field in the bright zone with minimum radiation into the dark zone. Both sound zones are in a plane levelled with the listener's ears.

Here, we use the SMM, SWMM, and WMM methods to design the loudspeaker driving signals, and the bright zone reproduction performance is measured through the Eigenmike recordings extracted modal-domain coefficients. Note that the effective observation range of the Eigenmike is sufficient to recover sound fields below 1 kHz within 0.15 m radius.

We define the modal-domain coefficients based reproduction error within the bright zone as[38]

$$\varsigma(k) = 10\log_{10}\left(\frac{||\mathbf{c}_{\text{des}} - \mathbf{c}||^2}{||\mathbf{c}_{\text{des}}||^2}\right), \tag{26}$$

where $\mathbf{c}_{\text{des}}$ and $\mathbf{c}$ denote column vectors of desired and measured bright zone modal coefficients, respectively.

The conventional method SMM was compared with the proposed weighting methods, SWMM and WMM. The results with and without ATF calibration are plotted across frequency range [200, 1000] Hz and distance range [0.2, 0.6] m in Figs. 10–12, respectively.

With the actual ATF knowledge, the AC result of SMM is lower than that of SWMM and WMM for almost all considered frequencies in the semi-anechoic experimental environment as shown in Figs. 10(a) and 10(b), except at the frequency of 500 Hz. This abnormal result may be caused by the ATF measurement error due to loudspeaker/microphone positioning errors, and sensor-self noise, etc. WMM demonstrates the lowest bright zone reproduction error over the entire frequency range as shown in Figs. 10(c) and 10(d). In addition, the conventional SMM has the lowest AC when the distance is larger than 0.3 m, corresponding to the optimal control points of SMM which are around the global origin. The experimental results are consistent with those from simulations as shown in Fig. 6.

Compared to the experimental results in the semi-anechoic environment, room reverberation degrades the overall sound field reproduction performance, such as around 3–7 dB acoustic contrast loss and 2–5 dB error loss as shown in Figs. 10 and 11. The proposed WMM method achieves the better AC results and the lowest reproduction

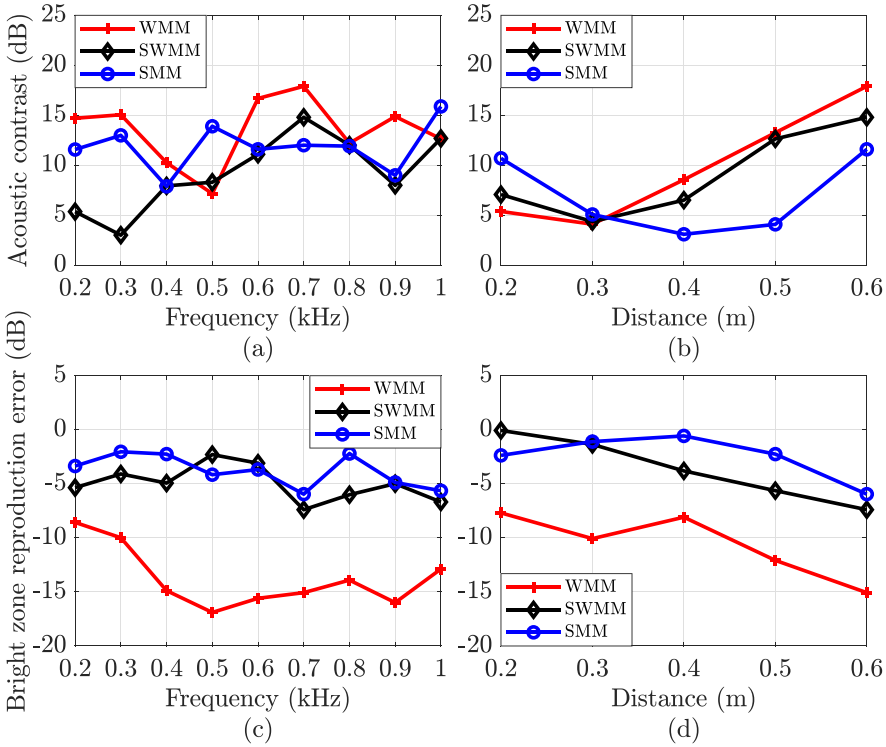

FIG. 10. (Color online) Semi-anechoic experimental results with ATF calibration: (a) and (c) are the acoustic contrast and bright zone reproduction error results varying with frequency when the distance between the local origin and global origin is fixed at 0.6 m. (b) and (d) are the acoustic contrast and bright zone reproduction results varying with distance when the frequency is fixed at 700 Hz.

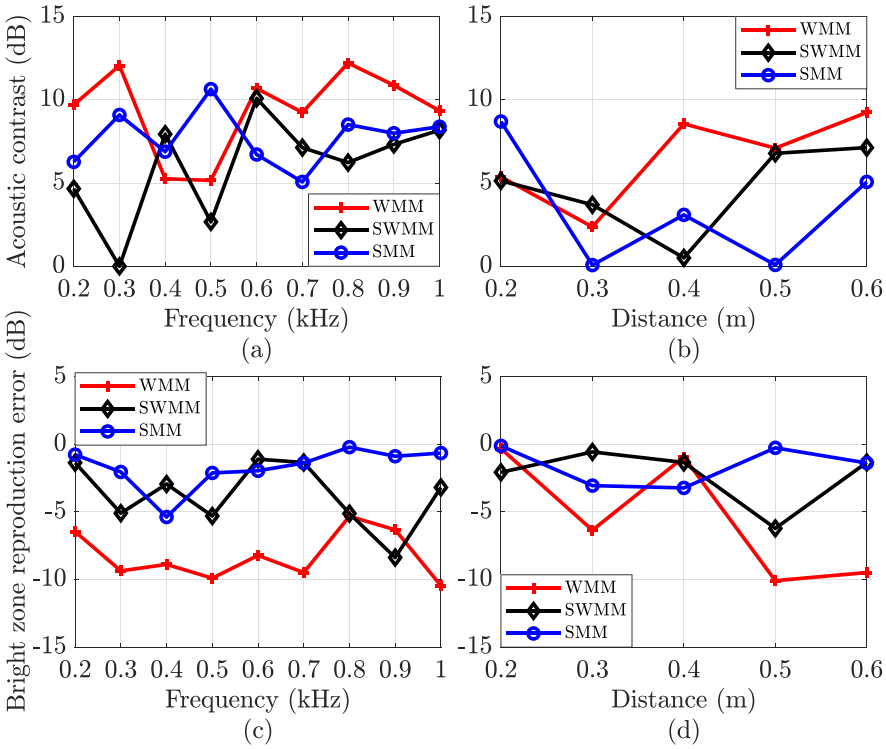

FIG. 11. (Color online) Reverberant experimental results with ATF calibration: (a) and (c) are the acoustic contrast and bright zone reproduction error results varying with frequency when the distance between the local origin and global origin is fixed at 0.6 m. (b) and (d) are the acoustic contrast and bright zone reproduction results varying with distance when the frequency is fixed at 700 Hz.

error in the reverberant acoustic environments. In addition, comparing the AC performance of these three methods, the WMM and SWMM approaches perform especially better when the distance between global and local region is larger than 0.33 m, as shown in Figs. 10(b) and 11(b). The results also demonstrate that the suppression of reverberation is very important for sound field reproduction.

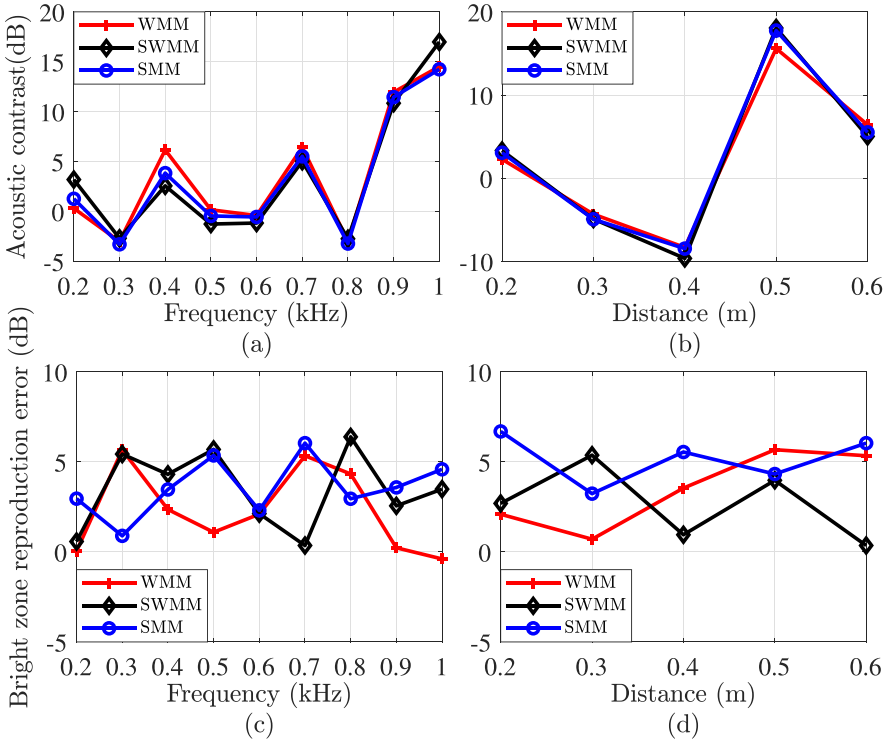

FIG. 12. (Color online) Reverberant experimental results without ATF calibration: (a) and (c) are the acoustic contrast and bright zone reproduction error results varying with frequency when the distance between the local origin and global origin is fixed at 0.6 m. (b) and (d) are the acoustic contrast and bright zone reproduction results varying with distance when the frequency is fixed at 700 Hz.

As shown in Figs. 12(a) and 12(b), without the actual ATF knowledge, i.e., the ATF of the loudspeaker is modelled as 3D free-field Green's function, these three methods have roughly the same AC performance and at some frequency points, the AC results are quite poor (below 0 dB). The bright zone reproduction error is also much greater comparing to the results in Fig. 11. As expected, the actual ATF knowledge is required in the mode-matching based method to get the desired results, the same requirement as in multi-point control based reproduction.

VI. CONCLUSION

In this work, the WMM approach was proposed for 2.5D multizone reproduction. The method is based on weighting the Bessel-spherical harmonic modes over the entire global control region. A detailed analysis of the weighting function was provided to provide insights and superiority of the proposed method. In comparison with the conventional approach, the advantages of the proposed weighting approach were confirmed by simulations-based and hardware-based experiments. The experimental results further showed that the proposed method can achieve better reproduction performance in both the semi-anechoic and reverberant environments.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (NSFC) funding scheme under Project No. 61671380. The author would like to thank Ms. J. Xie for her helpful comments.

APPENDIX A: SEMIDEFINITE RELAXATION

The first step for deriving an SDR of Eq. (6) is to use the cyclic property of the trace

$$\mathbf{x}^T\mathbf{A}_\mathrm{b}\mathbf{x} = Tr(\mathbf{x}^T\mathbf{A}_\mathrm{b}\mathbf{x}) = Tr(\mathbf{A}_\mathrm{b}\mathbf{x}\mathbf{x}^T). \tag{A1}$$

Both the objective function and constraints in Eq. (6) are linear in the matrix $\mathbf{x}\mathbf{x}^T$. By introducing a variable $\mathbf{X} = \mathbf{x}\mathbf{x}^T$, we have

$$\mathbf{x}^T\mathbf{A}_\mathrm{b}\mathbf{x} = Tr(\mathbf{A}_\mathrm{b}\mathbf{X}), \tag{A2}$$

where $\mathbf{X}$ is equivalent to a rank one symmetric positive semidefinite matrix, that is

$$\mathbf{X} \succcurlyeq 0, \quad rank(\mathbf{X}) \leq 1, \tag{A3}$$

where $\mathbf{X} \succcurlyeq 0$ denotes that $\mathbf{X}$ is positive semidefinite.

By using the result from Eq. (A2), we obtain the following equivalent formulation of Eq. (6):

$$\min_{\mathbf{X}} \quad Tr(\mathbf{A}_\mathrm{b}\mathbf{X}), \tag{A4}$$

$$\text{subject to} \quad Tr(\mathbf{A}_\mathrm{d}\mathbf{X}) \leq e_\mathrm{d}, \tag{A4a}$$

$$Tr(\mathbf{X}) \leq e_\mathrm{g}, \tag{A4b}$$

$$\mathbf{X} \succcurlyeq 0, \quad rank(\mathbf{X}) \leq 1, \tag{A4c}$$

$$t^2 = 1. \tag{A4d}$$

The results show that the only difficult constraint in Eq. (6) is the rank one constraint $rank(\mathbf{X}) \leq 1$, which is nonconvex (the objective and all other constraints are convex in $\mathbf{X}$). Therefore, we drop the rank one constraint to obtain the following relaxed from of Eq. (6) according to Ref. 40, that is

$$\min_{\mathbf{X}} \quad Tr(\mathbf{A}_\mathrm{b}\mathbf{X}), \tag{A5}$$

$$\text{subject to} \quad Tr(\mathbf{A}_\mathrm{d}\mathbf{X}) \leq e_\mathrm{d}, \tag{A5a}$$

$$Tr(\mathbf{X}) \leq e_\mathrm{g}, \tag{A5b}$$

$$\mathbf{X} \succcurlyeq 0, \tag{A5c}$$

$$t^2 = 1. \tag{A5d}$$

In this way, we transform the original complex-valued QCQP problem as an SDR.

APPENDIX B: ATF RECORDING USING HIGHER ORDER MICROPHONE

In our experiment, the loudspeakers are located completely outside of the control region, and thus the ATF measurements can be viewed as a spatial interior recording problem.[41,42] We adopt the method used in Ref. 38 to capture the sound pressure produced by the loudspeaker. A single Eigenmike is moved along a horizontal circle (assuming stationary conditions) to obtain the local sound pressure at different observation positions.

Then, we combine multiple local measurements and transform them to the global result using the spherical Bessel function addition theorem.[28] That is, one position sound field coefficient $\eta_{\nu\mu}$ can be related to the global sound field coefficients ζ_{vu} as follows:

$$\sum_{u=-N_0}^{N_0}\sum_{v=|u|}^{N_0}\zeta_{vu}(k)H_{\nu\nu}^{u\mu}(\mathbf{o}_p,k) = \eta_{\nu\mu}(k), \tag{B1}$$

where $\mathbf{o}_p$ is the vector that points from the global origin to the origin of the pth observation position, and

$$H_{\nu\nu}^{u\mu}(\mathbf{o}_p,k) = 4\pi i^{\nu-v}W_1W_2\sum_{l=0}^{\infty}i^l(-1)^{2u-\mu}j_l(k||\mathbf{o}_p||) \times \overline{Y_{l(\mu\bar{u})}}(\theta_{\mathbf{o}_p},\phi_{\mathbf{o}_p})\sqrt{\frac{(2v+1)(2\nu+1)(2l+1)}{4\pi}}, \tag{B2}$$

with

$$W_1 = \begin{pmatrix} v & \nu & l \\ 0 & 0 & 0 \end{pmatrix}, \quad W_2 = \begin{pmatrix} v & \nu & l \\ u & -\mu & (\mu-u) \end{pmatrix}, \tag{B3}$$

denoting the Wigner $3-j$ symbol.

Using the matrix-vector notation, we can represent Eq. (B1) as

$$\mathbf{Hg} = \mathbf{y}, \tag{B4}$$

where $\mathbf{g} = [\zeta_0, \ldots, \zeta_{\nu^2+\nu+u+1}, \ldots, \zeta_{(N_0+1)^2}]^T$, $\mathbf{y} = [\eta_0, \ldots, \eta_{1+\nu^2+\nu+\mu}, \ldots, \eta_{(N_p+1)^2}]^T$, and $\mathbf{H}$ is the $(N_p+1)^2 \times (N_0+1)^2$ translation matrix. The mode number upper limit of the local and global observation region are $N_p = \lceil ekr_p/2 \rceil$ and $N_0 = \lceil ekr_0/2 \rceil$, respectively.

Because $\mathbf{H}$ and $\mathbf{H}^H$ are one-sided inverse,[28] that is

$$\mathbf{HH}^H = \mathbf{I}. \tag{B5}$$

Thus, we have the following relationship:

$$\mathbf{H}^H\mathbf{y} = \mathbf{g}. \tag{B6}$$

Comparing to Eq. (B4), we can obtain the global coefficients using Eq. (B6) directly.

[1]T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," IEEE Signal Proc. Mag. **32**(2), 81–91 (2015).

[2]W. Zhang, P. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: A review of spatial audio recording and reproduction," Appl. Sci. **7**(5), 532 (2017).

[3]J. Cheer and S. J. Elliott, "Design and implementation of a personal audio system in a car cabin," in *Proceedings of the International Conference on Acoustics (ICA)*, Vancouver, Canada (May 26–31, 2013), p. 055009.

[4]S. J. Elliott and M. Jones, "An active headrest for personal audio," J. Acoust. Soc. Am. **119**(5), 2702–2709 (2006).

[5]H. So and J. W. Choi, "Subband optimization and filtering technique for practical personal audio systems," in *Proceedings of the IEEE ICASSP*, Brighton, UK (May 12–17, 2019), pp. 8494–8498.

[6]M. Buerger, C. Hofmann, and W. Kellermann, "Broadband multizone sound rendering by jointly optimizing the sound pressure and particle velocity," J. Acoust. Soc. Am. **143**(3), 1477–1490 (2018).

[7]J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," J. Acoust. Soc. Am. **131**(6), 4518–4525 (2012).

[8]J. W. Choi and Y. H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," J. Acoust. Soc. Am. **111**(4), 1695–1700 (2002).

[9]Y. Cai, M. Wu, L. Liu, and J. Yang, "Time-domain acoustic contrast control design with response differential constraint in personal audio systems," J. Acoust. Soc. Am. **135**(6), EL252–EL257 (2014).

[10]Q. Feng, F. Yang, and J. Yang, "Time-domain sound field reproduction using the group lasso," J. Acoust. Soc. Am. **143**(2), EL55–EL60 (2018).

[11]M. F. S. Gálvez, S. J. Elliott, and J. Cheer, "Time domain optimization of filters used in a loudspeaker array for personal audio," IEEE Trans. Audio Speech Lang. Process. **23**(11), 1869–1878 (2015).

[12]Z. Han, M. Wu, Q. Zhu, and J. Yang, "Two-dimensional multizone sound field reproduction using a wave-domain method," J. Acoust. Soc. Am. **144**(3), EL185–EL190 (2018).

[13]M. A. Poletti and F. M. Fazi, "An approach to generating two zones of silence with application to personal sound systems," J. Acoust. Soc. Am. **137**(2), 598–605 (2015).

[14]Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," IEEE Trans. Audio Speech Lang. Process. **19**(6), 1711–1720 (2011).

[15]W. Jin and W. B. Kleijn, "Theory and design of multizone soundfield reproduction using sparse methods," IEEE Trans. Audio Speech Lang. Process. **23**(12), 2343–2355 (2015).

[16]M. A. Poletti and F. M. Fazi, "Generation of half-space sound fields with application to personal sound systems," J. Acoust. Soc. Am. **139**(3), 1294–1302 (2016).

[17]N. Ueno, S. Koyama, and H. Saruwatari, "Three-dimensional sound field reproduction based on weighted mode-matching method," IEEE Trans. Audio Speech Lang. Process. **27**(12), 1852–1867 (2019).

[18]J. Ahrens and S. Spors, "An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions," Acta Acust. united Ac. **94**(6), 988–999 (2008).

[19]S. Spors, V. Kuscher, and J. Ahrens, "Efficient realization of model-based rendering for 2.5-dimensional near-field compensated higher order ambisonics," in *Proceedings of the IEEE WASPAA*, New Paltz, NY (October 16–19, 2011), pp. 61–64.

[20]F. Winter, J. Ahrens, and S. Spors, "On analytic methods for 2.5-D local sound field synthesis using circular distributions of secondary sources," IEEE Trans. Audio Speech Lang. Process. **24**(5), 914–926 (2016).

[21]W. Zhang and T. D. Abhayapala, "2.5 D sound field reproduction in higher order ambisonics," in *Proceedings of IEEE IWAENC*, Juan-les-Pins, France (September 8–11, 2014), pp. 342–346.

[22]T. Okamoto, "2.5D higher order ambisonics for a sound field described by angular spectrum coefficients," in *Proceedings of the IEEE ICASSP*, Shanghai, China (March 20–25, 2016), pp. 326–330.

[23]T. Okamoto, "Analytical approach to 2.5D sound field control using a circular double-layer array of fixed-directivity loudspeakers," in *Proceeding of the IEEE ICASSP*, New Orleans, LA (March 5–9, 2017), pp. 91–95.

[24]T. Okamoto, "Horizontal 3D sound field recording and 2.5D synthesis with omni-directional circular arrays," in *Proceedings of the IEEE ICASSP*, Brighton, UK (May 12–17, 2019), pp. 960–964.

[25]W. Zhang, J. Zhang, T. D. Abhayapala, and L. Zhang, "2.5 D multizone reproduction using weighted mode matching," in *Proceedings of the IEEE ICASSP*, Calgary, Canada (April 15–20, 2018), pp. 476-480.

[26]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, New York, 1999), Chap. 4, pp. 115–148.

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

[28]P. A. Martin, *Multiple Scattering: Interaction of Time-Harmonic Waves With N Obstacles* (Cambridge University Press, Cambridge, UK, 2006), Chaps. 2–3, pp. 29–121.

[29]W. Zhang, T. D. Abhayapala, T. Betlehem, and F. M. Fazi, "Analysis and control of multi-zone sound field reproduction using modal-domain approach," J. Acoust. Soc. Am. **140**(3), 2134–2144 (2016).

[30]T. Betlehem and P. D. Teal, "A constrained optimization approach for multi-zone surround sound," in *Proceedings of the IEEE ICASSP*, Prague, Czech Republic (May 22–27, 2011), pp. 437–440.

[31]Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," IEEE Signal Proc. Mag. **27**(3), 20–34 (2010).

[32]T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," J. Acoust. Soc. Am. **117**(4), 2100–2111 (2005).

[33]T. D. Abhayapala and A. Gupta, "Spherical harmonic analysis of wavefields using multiple circular sensor arrays," IEEE Trans. Audio Speech Lang. Process. **18**(6), 1655–1666 (2010).

[34]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (National Bureau of Standards, New York, 1970), Chap. 8, p. 334.

[35]M. A. Poletti, T. Betlehem, and T. D. Abhayapala, "Analysis of 2D sound reproduction with fixed-directivity loudspeakers," in *Proceedings IEEE ICASSP*, Kyoto, Japan (March 25–30, 2012), pp. 377–380.

[36]J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**(4), 943–950 (1979).

[37]A. Farina, A. Capra, L. Chiesi, and L. Scopece, "A spherical microphone array for synthesizing virtual directive microphones in live broadcasting and in post production," in *Proceedings of the 40th Audio Engineering Society (AES) Conference*, Tokyo, Japan (October 8–10, 2010), pp. 1–11.

[38]Y. Hu, P. N. Samarasinghe, G. Dickins, and T. D. Abhayapala, "Modeling the interior response of real loudspeakers with finite measurements," in *Proceeding of the IEEE IWAENC*, Tokyo, Japan (September 17–20, 2018), pp. 16–20.

https://doi.org/10.1121/10.0000797

[39]M. Binelli, A. Venturi, A. Amendola, and A. Farina, "Experimental analysis of spatial properties of the sound field inside a car employing a spherical microphone array," in *Proceedings of the 130th Audio Engineering Society (AES) Conference*, London, UK (May 13–16, 2011), pp. 1–8.

[40]Y. Huang and D. P. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," IEEE Trans. Signal Process. **58**(2), 664–678 (2009).

[41]P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "3D spatial soundfield recording over large regions," in *Proceedings of the IEEE IWAENC*, Aachen, Germany (September 4–6, 2012), pp. 1–4.

[42]P. N. Samarasinghe, T. D. Abhayapala, and M. A. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," IEEE Trans. Audio Speech Lang. Process. **22**(3), 647–658 (2014).