Sound field reproduction with a Cylindrical Loudspeaker Array using first order wall reflections

Natsuko Maeda, Filippo Maria Fazi, and Falk-Martin Hoffmann

Abstract—**Sound field reproduction is a promising technology for reproducing a desired sound field over a specific area. However, several practical challenges exist, such as the requirement for a large number of loudspeakers and the degradation of spatial reproduction accuracy due to room reflections. This work proposes a sound field reproduction method based on a cylindrical loudspeaker array that is used to both radiate sound directly towards the listener and take advantage of controlled first-order reflections from the room's walls. The sound field corresponding to one or more virtual sound sources is created by the combination of the direct and reflected sound at the listener position. As long as the lateral walls generate well-defined specular reflections, the simple model of the reverberant field with the proposed formulation enables this method to work without elaborate acoustical measurements. Two approaches with different concepts of controlling the array are compared and their mathematical equivalence is shown. To evaluate the performance, both simulations and experiments were conducted, showing a good agreement with theory.**

Index Terms—**Cylindrical Loudspeaker Array (CLA), wall reflections, image source method, mode matching.**

I. Introduction

SOUND field reproduction systems delivering immersive sound experiences in home environments have gained new relevance due to the rapid development and growth of streaming services. The introduction of object-based audio contents for rendering on systems such as Dolby Atmos, DTS:X, and Sony's 360 Reality Audio accelerate the development of the technology that renders the spatial information describing the sound scene.

Several technologies to reproduce the sound field using multiple loudspeakers for generating spatial audio experience over an extended listening area have been proposed over the past few decades, such as Wave field synthesis (WFS) and Higher Order Ambisonics (HOA) [1]–[5]. However, most of these techniques are derived on the basis of the use of loudspeakers, which can be regarded as monopole sound sources in the free sound field with no adjacent reflecting surfaces. These assumptions are not valid in practice and cause impaired spatial reproduction.

There are two major approaches to reducing the influence of the room on the reproduced sound field. One is active room compensation that requires the knowledge of acoustic transfer function between the loudspeaker and multiple control points. The traditional approach has been to equalize the transfer functions over multiple control points using least-squares techniques, which has been widely introduced to consumer audio products [6]–[9]. However, these approaches may result in poor equalization between the control points [10]. Other recent approaches for active room compensation are based on WFS [11]–[13] or HOA [14]–[16]. These approaches can control the sound field over an extended contiguous area but may require considerable processing for the implementation of a digital compensation system.

The alternative approach is to use directional sound sources for limiting room reflections or exploiting the reflections to create the desired sound field. Two types of directional sound sources are proposed such as fixed directivity loudspeakers [17] and variable directivity loudspeakers [18] to limit room reflections. Fixed directivity loudspeakers such as cardioids can reduce the sound wave radiated outside the area the array encompasses [17]. The approach with variable directivity loudspeakers is based on the Kirchoff-Helmholtz integral and allows the complete cancellation of the externally radiating sound [18]. However, these approaches require numerous transducers to control over a wide range of frequencies.

The approach using steerable directional sound sources to exploit room reflections is one of the promising ways that can reduce the number of sources/loudspeakers when compared with previous methods [19]–[24]. Many of these systems, however, require measuring the acoustics transfer functions between each directional sound source and each control point or sound field mode at the listening point using a microphone array. These acoustic transfer functions include higher-order reflections from walls and both reflected and scattered sounds from furniture in a room, which may reduce the system robustness due to the time variability of said transfer functions.

Perceptual tests of the WFS in a reverberant sound field by Erbes et al. [25] show that if the speakers are placed closely together and the reverberation time is not too long, the effect of reflected sound can be ignored due to the precedence effect. While this is an interesting result, it suggests that there are challenges in living room environments where the reverberation time is long and many speakers cannot be placed.

In this work, we employ the concept of sound field reproduction with sound sources with adjustable directivity and using specular reflections from walls that can reduce the number of speakers to be installed in a room. To achieve a robust sound field reproduction with limited knowledge of the room, a new formulation is introduced by simplifying the model with some approximations. The first simplification is using only first-order reflections, which are strong enough to reproduce the desired sound field. This simplification is expected to be effective because only early reflections affect the directional perception [26]. Hence, in contrast to some of the existing approaches [14], [21], the proposed method does not rely on a complete model of the reproduced sound field,

but is based on a simplified model consisting of the direct sound and several image sources associated to wall reflections.

The second simplification relies on the assumption that the distance between the sound source and the listening position is sufficiently large, with respect to the wavelength of the sound to be reproduced. This allows for a simplified formulation. Two mode-matching approaches with different concepts are compared mathematically, one is a direct mode-matching approach with the simplified model described above, and the other is a beam steering approach similar to the concept of Yamaha's YSP series and previous work [22], [27]. For this latter approach, the systems steers a beam towards the listener and other sound beams towards the room's walls. These beams are delayed in such a way that the direct and reflected waves reach the listener simultaneously, and their relative amplitude generated the perception of a virtual sound source, similarly to what happens with amplitude panning. One of the contributions of this work is the proof that the presumably more intuitive beam steering approach outlined above is mathematically equivalent to the more complex direct mode-matching approach, which is mathematically more involved and does not explicitly include beamforming in its formulation. This result, which is a consequence of the simplified sound field model adopted in this work, sheds some light on the physical phenomena underpinning the proposed method and allows for a better understanding of its limitations.

Experiments are conducted to validate the effectiveness of the proposed approach in real situations. We discuss how the loudspeakers characteristics such as frequency response and dynamic range and loudspeaker positioning errors affect the accuracy of the sound field.

The article is organized as follows. Section II describes the sound field radiated from a Cylindrical Loudspeaker Array (CLA). In Section III, the room reflections model is introduced not only for a monopole sound source but also for a directional sound source. We propose two approaches, namely direct mode matching and beam steering approach, and demonstrate their mathematical equivalence in Section IV. The obtained experimental results are described and discussed in Section V, which is followed by some concluding considerations in Section VI.

II. Loudspeaker array model

This section describes the mathematical model for the sound field radiated from a CLA. There are several articles and books that deal with sound field reproduction using a CLA [28]–[35]. In this work, we derive a simplified formula to expand the model with room reflections later. The geometrical description of this model is based on the cylindrical coordinate system as shown in Figure 1. All considerations presented in the following are in the frequency domain, and the dependency on $e^{-i\omega t}$ and the parameter ω are omitted in all expressions for reasons of brevity.
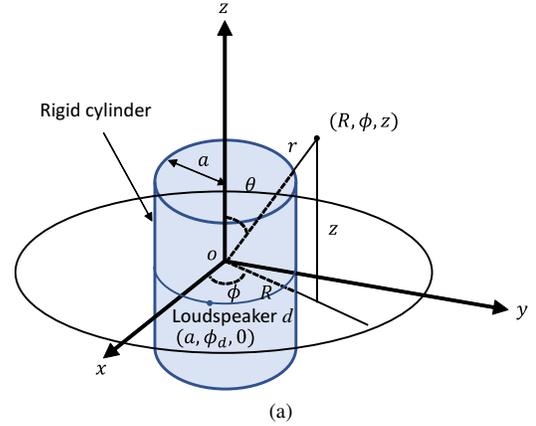
Fig. 1: Reference geometry for the Cylindrical Loudspeaker Array.

A. Sound field model for Cylindrical waves

The sound field generated by a cylinder vibrating in an arbitrary fashion is given by the expression [35]

$$
\begin{aligned}
&p(R, \phi, z) \\
&= \sum_{\mu=-\infty}^{\infty} e^{i\mu\phi} \int_{-\infty}^{\infty} i\rho c k \frac{H_\mu(k_r R)}{k_r H'_\mu(k_r a)} \frac{1}{2\pi} e^{i k_z z} V_\mu(a, k_z)\, dk_z,
\end{aligned}
\tag{1}
$$

where ρ is the density of air, c is the sound speed of air, k is the wave number, $k_r = \sqrt{k^2 - k_z^2}$, and $H_\mu(\cdot)$ and $H'_\mu(\cdot)$ are the Hankel function of the first kind of order μ and its derivative with respect to the function argument, respectively. $V_\mu(a, k_z)$ is the two-dimensional Fourier transform in ϕ and z of the radial velocity $v(a, \phi, z)$ at $r = a$, which could also be referred to as the helical wave spectrum of the radial velocity and is given by

$$
\begin{aligned}
V_\mu(a, k_z) &= \frac{1}{2\pi} \int_0^{2\pi} d\phi \int_{-\infty}^{\infty} v(a, \phi, z) e^{-i\mu\phi} e^{-i k_z z} dz \\
&= F_\phi F_z [v(a, \phi, z)].
\end{aligned}
\tag{2}
$$

Equations (1) and (2) show the relationship between the radial velocity on the cylindrical baffle and an arbitrary radiated pressure field.

B. Driving function of the loudspeakers on a rigid cylinder

Consider a cylindrical rigid baffle of radius a and infinite extent in z, and assume the loudspeaker on the baffle as a monopole sound source. The corresponding velocity profile of the dth loudspeaker on the intersection of the cylinder surface with the plane $z = 0$ is given by

$$
v(a, \phi, z) = v_d \delta(z) \delta(\phi - \phi_d).
\tag{3}
$$

For an array of D monopole sound sources at angle ϕ_d with $d = 1 \dots D$, the sum of all loudspeaker helical spectra is

$$
V_\mu(a, k_z) = \frac{1}{2\pi} \sum_{d=1}^{D} v_d e^{-i\mu\phi_d},
\tag{4}
$$

which corresponds to the spatial *DFT* of the velocity of the loudspeakers v_d if $\phi_d = \frac{2\pi d}{D}$. As the coefficients V_μ do not depend on k_z or on the cylinder radius a, we will omit these variables in the notation of V_μ. Since an electrodynamic loudspeaker operates largely in its mass-controlled region, the electric driving signals are proportional to the mass acceleration q_d, given by

$$q_d = \rho \dot{v}_d = -i\omega \rho v_d. \tag{5}$$

We therefore define the mass acceleration spectrum, hereafter referred to as the CLA angular spectrum or CLA modal coefficients, as

$$Q_\mu = -i\rho\omega V_\mu. \tag{6}$$

Replacing V_μ with Q_μ yields the expression

$$p(R,\phi,z) = \sum_{\mu=-\infty}^{\infty} e^{i\mu\phi} \left[-\int_{-\infty}^{\infty} \frac{H_\mu(k_r R)}{2\pi k_r H'_\mu(k_r a)} e^{ik_z z} dk_z \right] Q_\mu. \tag{7}$$

C. Far field approximation

When r is very large compared to the wavelength, the Hankel function of the numerator in Equation (7) can be replaced with the far field approximation (see Appendix)

$$H_\mu(x) \approx \sqrt{\frac{2}{\pi x}} e^{i(x-\mu\pi/2-\pi/4)}. \tag{8}$$

Since the integration over k_z in Equation (7) is not trivial, we apply the Stationary Phase Approximation [35] to simplify it and obtain the following formula for the sound field reproduced by the CLA:

$$p(r,\theta,\phi) = G(r)D(\theta,\phi), \tag{9}$$

where $R = r\sin\theta$ and $z = r\cos\theta$,

$$G(r) = \frac{e^{ikr}}{4\pi r} \tag{10}$$

is the free field Green's function and $D(\theta,\phi)$ is the directivity function, given by

$$D(\theta,\phi) = \sum_{\mu=-\infty}^{\infty} \underbrace{\frac{4(-i)^{\mu-1}}{k\sin\theta H'_\mu(ka\sin\theta)}}_{R_\mu(\theta)} e^{i\mu\phi} \, Q_\mu(a, k\cos\theta). \tag{11}$$

Note that, due to our choice of the velocity profile for our loudspeaker defined in Equation (3), the stationary phase approximation for the radiated pressure in Equation (9) shows a radial decay with $1/r$. This leads to our individual source being similar to a point source on a cylindrical baffle, since it does not assume the $1/\sqrt{r}$ decay inherent to the large argument approximation of Hankel functions [35], which are the radial functions of the expansion in Equation (1).
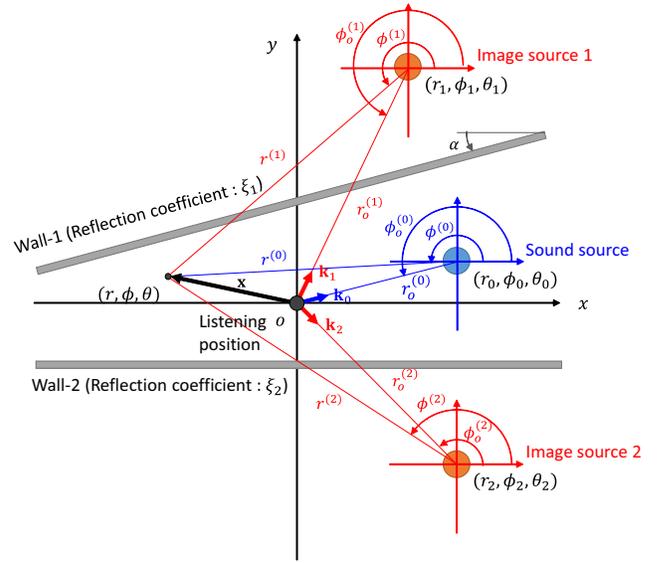
Fig. 2: Reference geometry, with the Cylindrical Loudspeaker Array and two vertical walls.

III. REFLECTION MODEL

In this section, the image source method is introduced and used to model wall reflections. Figure 2 shows the reference geometry and the definition of each parameter. Note that the listening position o is chosen as the origin of the absolute reference system.

The image source method is based on geometrical acoustics and is widely used to model early reflections [36]–[41]. The total sound pressure at point (r,θ,ϕ) in the vicinity of the listening position can be modeled by summing the fields generated by the actual sound source and by all the audible image sources

$$p(r,\theta,\phi) = \sum_{l=0}^{L} p_l(r,\theta,\phi), \tag{12}$$

where L is the number of image sources, $p_0(r,\theta,\phi)$ is the sound field created by the real sound source in the free field (the direct sound), and $p_{l\geq 1}(r,\theta,\phi)$ is the sound field created by the lth image source.

To further motivate the use of this model, Figure 3 shows the impulse response measured in a typical listening room of dimensions $4.2\,\text{m} \times 6.3\,\text{m} \times 2.5\,\text{m}$ and of T60 approximately $0.36\,\text{s}$ using 2-way active speaker as the sound source, and simulation results under the same condition. The first order reflections of the simulation are well-matched with the measured ones, while measured second or higher order reflections are hard to distinguish.

Room reflections of a directional sound source can be modeled using an extension of the image source method. Since the sound field generated by a CLA is expressed as the product of the free field Green's function and the directivity function, as shown in Section II, each reflection generated by a directional sound source and any vertical wall can be modelled as

$$p_l(r,\theta,\phi) = \xi_l G(r^{(l)}) D\left(\theta^{(l)}, T_l[\phi^{(l)}]\right), \tag{13}$$
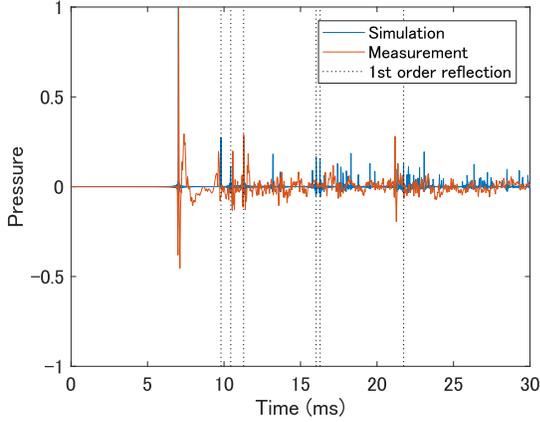
Fig. 3: Typical impulse response of listening room $4.2\,\mathrm{m} \times 6.3\,\mathrm{m} \times 2.5\,\mathrm{m}$.

where $\left(r^{(l)}, \theta^{(l)}, \phi^{(l)}\right)$ are the coordinates of a point (r, θ, ϕ) expressed with respect to the reference system consistent with the location of the lth image source (Figure 2) and ξ_l is the accumulated reflection coefficient resulting from multiple reflection off the walls corresponding to the lth image source. T_l is the mirroring operators for the walls, which ensures that the directivity pattern of the image source for a first order reflection is given by the directivity function of the primary source but mirrored with respect to the given wall [20]. Multiple mirroring operations are required for image sources of higher order. The 0th image source corresponds to the direct sound source, with $\xi_0 = 1$ and $T_0[\phi^{(0)}] = \phi^{(0)}$.

If there are only two walls and they are parallel to the xz-plane, the reflected sound created by the lth first-order image source is

$$p_{l\geq 1}(r, \theta, \phi) = \xi_l\, G(r^{(l)})\, D(\theta^{(l)}, -\phi^{(l)}). \qquad (14)$$

In the case that a given vertical wall is not parallel to the xz-plane but is rotated by an angle α with respect to that plane, as shown in Figure 2, the mirroring operator becomes more complex, with $T_l[\phi^{(l)}] = -\phi^{(l)} + 2\alpha$.

Assuming that the real and image directional sound sources are far away from the listening position and that the observation point (r, θ, ϕ) is close to the listening position, the free field Green's function in Equation (13) can be locally approximated with a plane wave as

$$G(r^{(l)}) \approx G(r_o^{(l)})e^{-ik_l \cdot x}, \qquad (15)$$

where $x = [r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta]$, k_l is the wave number vector, with length k and pointing in the direction of the lth image source, and $r_o^{(l)}$ is the distance of the lth source from the listening position, as shown in Figure 2. Also, the directivity function of Equation (11) can be approximated as

$$D(\theta^{(l)}, \phi^{(l)}) \approx D(\theta_o^{(l)}, \phi_o^{(l)}), \qquad (16)$$

where the relative angular coordinates of the observation point $\theta^{(l)}, \phi^{(l)}$ have been substituted with the relative angular coordinates of the listener position $\theta_o^{(l)}, \phi_o^{(l)}$, under the assumption that $r^{(l)} \gg r$. Note that, based on the reference geometry depicted in Figure 2, $\theta_o^{(l)} = \pi - \theta_l$ and $\phi_o^{(l)} = \pi + \phi_l$.

In the light of these simplifications, the sound field reproduced by the CLA is given by

$$p(r, \theta, \phi) = \sum_{l=0}^{L} \xi_l\, G(r_o^{(l)})\, D\left(\theta_o^{(l)}, T_l[\phi_o^{(l)}]\right)\, e^{-ik_l \cdot x}. \quad (17)$$

This means that the sound field in the vicinity of the listener is represented by a superposition of plane waves $e^{-ik_l \cdot x}$, each of which corresponds to either the direct sound or to an image source, and is scaled by the wall reflection coefficient ξ_l, by the attenuation and delay factor $G(r_o^{(l)})$, and by the associated value of the CLA directivity $D\left(\theta_o^{(l)}, T_l[\phi_o^{(l)}]\right)$.

IV. Mode matching with a CLA using wall reflections

In this section, sound field reproduction with a CLA using first-order reflections from two walls is discussed. Figure 3 indicates that higher-order reflections (second or higher order in the case considered), with long paths, may be difficult to control effectively for reproducing the desired sound field due to sound absorption and diffusion by obstacles, as well as the significant influence of temperature change that affects the phase. For this reason, we will neglect higher-order reflections and consider only first-order reflections ($L = 2$) from the two walls. This will of course have an effect on the reproduced sound field but it is expected that higher-order reflections will give little contribution to sound localisation cues, since these reflections will reach the listener at a later time than the direct sound and first-order reflections (which will be time-aligned to the direct sound, as shown later). Nevertheless, the theoretical arguments presented in this section can be straightforwardly extended to a greater reflection order or to a larger number of walls.

Two approaches are proposed: the mode matching and the beam steering approach. It will be shown later that the two approaches are in fact equivalent.

A. Direct mode matching approach

The objective of this method is to match, in a least-squares sense, the modal coefficients of the sound field reproduced by the loudspeaker array with those of the desired sound field. By "modes" we refer to a set of basis functions used to represent the sound field. When the modes are spherical harmonic functions, the area where accurate sound field control can be achieved is a sphere, whose size is determined by the frequency and the number (i.e. highest order) of the controlled modes. Mode matching is directly related to HOA as shown for example in references [2] [4]. In the proposed approach, the modes of the reproduced sound field are expressed as a function of the modal coefficients of the directivity function of the CLA. Hence the mode matching approach provides the directivity function modal coefficients that generate the desired modal coefficients of reproduced field.

The target sound field is chosen to be a plane wave arriving from angles ϕ_D, and associated wave vector $\boldsymbol{k_D}$. Its spherical harmonic expansion is [35]

$$
\begin{aligned}
p_D(r,\theta,\phi) &= e^{-i\boldsymbol{k_D}\cdot\boldsymbol{x}} \\
&= 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (-i)^n j_n(kr) Y_n^m(\theta,\phi) Y_n^m(\theta_D,\phi_D)^*,
\end{aligned}
\tag{18}
$$

where $j_n(\cdot)$ denotes the spherical Bessel function of order n. The spherical harmonics are defined as [35]

$$
Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}}\sqrt{\frac{(n-m)!}{(n+m)!}} P_{nm}(\cos\theta) e^{im\phi}, \tag{19}
$$

where $P_n^m(\cdot)$ is the associated Legendre function. As discussed in Section III, the reproduced sound field generated by the direct sound and two first-order wall reflections is given by Equation (17), which combined with Equation (18) (with θ_l, ϕ_l instead of θ_D, ϕ_D) gives

$$
\begin{aligned}
p_R(r,\theta,\phi) =&4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (-i)^n j_n(kr) Y_n^m(\theta,\phi) \\
&\times \sum_{l=0}^{2} \xi_l G\left(r_o^{(l)}\right) D\left(\theta_o^{(l)}, T_l[\phi_o^{(l)}]\right) Y_n^m(\theta_l,\phi_l)^*.
\end{aligned}
\tag{20}
$$

The system of equations for the mode matching problem is obtained by combining Equations (11), (18), and (20), yielding

$$
\begin{aligned}
Y_n^m(\theta_D,\phi_D)^* =&\\
\sum_{\mu=-M}^{M} \sum_{l=0}^{2} &\xi_l G\left(r_o^{(l)}\right) R_\mu\left(\theta_o^{(l)}\right) Y_n^m(\theta_l,\phi_l)^* \; e^{i\mu T_l[\phi_o^{(l)}]} \, Q_\mu
\end{aligned}
\tag{21}
$$

for all n, m combinations. Note that, by design, the sum describing the directivity pattern of the array been truncated to the order $M \le (D-1)/2$, relying on the assumption that the array can generate an order-limited pattern only. This is possible as long as spatial aliasing can be ignored, which is true up to a given high-frequency limit, as demonstrated below and discussed in reference [42]. This can be written in matrix form as

$$
\boldsymbol{b} = \boldsymbol{\Psi}\boldsymbol{q}, \tag{22}
$$

where the vector $\boldsymbol{b}$ contains the modal coefficients $Y_n^m(\theta_D,\phi_D)^*$ of the desired sound field (up to a given truncation order N) and vector $\boldsymbol{q}$ the unknown modal coefficients Q_μ of the directional sound source. The elements of matrix $\boldsymbol{\Psi}$ relating the sound field modal coefficients with indices n, m to the CLA modal coefficient Q_μ is given by[1]

$$
\Psi_{nm,\mu} = \sum_{l=0}^{2} \xi_l \, G\left(r_o^{(l)}\right) R_\mu\left(\theta_o^{(l)}\right) Y_n^m(\theta_l,\phi_l)^* \, e^{i\mu T_l[\phi_o^{(l)}]}. \tag{23}
$$

[1]Note that the matrix element indexing used here is not rigorous, as suitable mapping between the coefficient indices n, m and μ and the indices of the elements of the matrix would be required. For example, $\Psi_{\zeta,\nu}$ with $\zeta = n^2 + n + m + 1$ and $\nu = \mu + M + 1$. This has, however, been omitted to simplify the notation.

This system of equations expresses the spherical harmonic coefficients of the reproduced sound field as a function of the CLA modal coefficients, which are the unknown of the problem.

In the case that $\theta_o^{(l)} = \pi/2$ for all sources and that the vertical walls are parallel, the associated Legendre function in the definition of the spherical harmonics as well as the terms under square root in Equation (19) appear identically on both sides of the mode matching equation and can therefore be removed. Hence, the modal coefficients of the desired sound field and the mode transfer matrix simplify as follows:

$$
\boldsymbol{b} = \left[e^{-iN\phi_D}, \cdots, e^{in\phi_D}, \cdots, e^{iN\phi_D} \right]^T, \tag{24}
$$

$$
\Psi_{n,\mu} = R_\mu(\pi/2) \sum_{l=0}^{2} \xi_l \, G\left(r_o^{(l)}\right) \; e^{-in\phi_l} \; e^{i\mu T_l[\phi_o^{(l)}]} \tag{25}
$$

This also reduces the number of non-identical rows of $\boldsymbol{\Psi}$ from $(N+1)^2$ to $2N+1$.

The CLA modal coefficients are obtained by minimizing the squared error between the modal coefficients of the sound field reproduced by the loudspeakers and those of the desired sound field. In the case of an under-determined system, i.e. when $\boldsymbol{\Psi}$ has more columns than linearly independent rows, the solution is given by

$$
\tilde{\boldsymbol{q}} = \boldsymbol{\Psi}^H [\boldsymbol{\Psi}\boldsymbol{\Psi}^H]^{-1}\boldsymbol{b} = \boldsymbol{\Psi}^\dagger \boldsymbol{b}. \tag{26}
$$

The proposed approach models the sound field as the superposition of three plane waves (direct sound and two wall reflections). Hence, even if a large number of directional sources is used, the rank of the matrix $\boldsymbol{\Psi}$ is three and therefore we cannot control more than three spherical harmonic coefficients of the reproduced field.

This method is related to that proposed by Betlehem, Poletti, and Abhayapala in [21] and [14], with two fundamental differences: firstly, the proposed method only controls the room's first order reflections and does not attempt to completely remove higher order reflections or to control late reverberation, other than trying to minimise the overall acoustic energy generated by the system. This avoids the robustness limitation typical of dereverberation systems. Secondly, the CLA acoustical transfer functions are not measured but are mathematically modelled by means of images sources. This enables the direct link to the beamforming approach discussed in the following subsection, which provides a greater understanding of the physics of the system and of its limitations.

B. Beam steering approach

The second approach is the beam steering method. In this case the system is split into two sub-systems: an Ambisonic decoder and a set of beamformers. Firstly, we assume that the reproduced sound field is generated by three plane waves. Their directions correspond to the angles of incidence from the actual, physical source (i.e the CLA) and the two controlled wall reflections. Furthermore, we assume that we can control these plane waves independently in terms of their magnitude and phase. We then create a first order Ambisonics decoder

that provides the three plane wave gains required for the reproduction of the desired field in the proximity of the listener, expressed in terms of spherical harmonics coefficients.

The resulting Ambisonics decoding equation is

$$\boldsymbol{b} = \boldsymbol{Y}\boldsymbol{w}, \tag{27}$$

where $\boldsymbol{b}$ is the modal coefficient of the desired sound field as shown in Equation (22), $\boldsymbol{w} = [w_0, w_1, w_2]^T$ is the amplitude of each synthesized plane wave and the 3D Ambisonics mode transfer matrix is

$$\boldsymbol{Y} = \begin{bmatrix} Y_0^0(\theta_0, \phi_0)^*, & Y_0^0(\theta_1, \phi_1)^*, & Y_0^0(\theta_2, \phi_2)^* \\ \vdots & \vdots & \vdots \\ Y_N^N(\theta_0, \phi_0)^*, & Y_N^N(\theta_1, \phi_1)^*, & Y_N^N(\theta_2, \phi_2)^* \end{bmatrix}. \tag{28}$$

We have assumed here that both the target plane wave and the three reproduced plane waves have zero phase at the listener position, hence we expect the weights $\boldsymbol{w}$ to be real-valued and therefore frequency-independent.

The second step is the creation of the beamformer. The synthesized total directivity function $D(\theta, \phi)$ is the linear combination of three directivity functions for the CLA, each of which aims to create a single plane wave at the listening position, arriving from the direction of either the real source or an image source, and scaled by the weight w_l created in the previous step. In practice, each directivity function creates a sound beam in the direction of the listener (for the real source) or of the image source and generates a null in the directivity pattern in the other two directions.

To that end, the vector $\boldsymbol{d}$ is defined, which specifies the value of the total directivity function in the three relevant directions, namely

$$\boldsymbol{d} = [D(\theta_0, \phi_0), \ D(\theta_1, T_1[\phi_1]), \ D(\theta_2, T_2[\phi_2])]^T. \tag{29}$$

We then impose that each element of this vector satisfies the following equation

$$G(r_o^{(l)})\xi_l d_l = w_l, \quad l = 0, 1, 2. \tag{30}$$

This yields the matrix equation

$$\boldsymbol{w} = \boldsymbol{G}\boldsymbol{d} = \boldsymbol{G}\boldsymbol{\Delta}\boldsymbol{q}, \tag{31}$$

where matrix $\boldsymbol{\Delta}$ is such that $\boldsymbol{d} = \boldsymbol{\Delta}\boldsymbol{q}$ and its terms are given by Equation (11) and

$$\boldsymbol{G} = \begin{bmatrix} G(r_o^{(0)}) & 0 & 0 \\ 0 & \xi_1 G(r_o^{(1)}) & 0 \\ 0 & 0 & \xi_2 G(r_o^{(2)}) \end{bmatrix} \tag{32}$$

The terms in matrix $\boldsymbol{G}$ are required to compensate for the propagation delay and attenuation caused by the wave propagation distance and by the wall reflections.

The Ambisonic decoding and the beamforming steps are combined by substituting Equation (31) into Equation (27), thus obtaining the following new equation

$$\boldsymbol{b} = \boldsymbol{Y}\boldsymbol{G}\boldsymbol{\Delta}\boldsymbol{q} = \boldsymbol{\Lambda}\boldsymbol{q}. \tag{33}$$

As in the mode-matching case, the above formula expresses the modal coefficients of the reproduced field, $\boldsymbol{b}$, as functions

of the CLA modal coefficients, $\boldsymbol{q}$. The elements of the mode transfer matrix $\boldsymbol{\Lambda}$ are

$$\Lambda_{nm,\mu} = \sum_{l=0}^{2} Y_n^m(\theta_l, \phi_l)^* \ \xi_l \ G\left(r_o^{(l)}\right) \ R_\mu\left(\theta_o^{(l)}\right) \ e^{i\mu T_l[\phi_o^{(l)}]}. \tag{34}$$

Comparing Equation (23) with (34) shows that $\boldsymbol{\Lambda} = \boldsymbol{\Psi}$ and that the beamforming approach is indeed equivalent to the direct mode-matching approach. It is interesting that also the mode-matching approach relies, in practice, on the creation of three (or more) sound beams, even if these are not explicitly included in the formulation of this method. This aspect will become very clear in the CLA radiation pattern simulations discussed in the next section.

Equation (33) also shows that, since $\boldsymbol{G}$ is a 3×3 matrix, the rank of $\boldsymbol{\Lambda}$ and of $\boldsymbol{\Psi}$ cannot be greater than 3, which also implies that only 3 degrees of freedom of the sound field can be independently controlled, as already discussed in the previous sub-section. This would of course be different if a greater number of reflections were considered in the problem.

V. EXPERIMENTS

To validate the performance of the proposed sound field reproduction method with a CLA and using first-order reflections described in Section IV, numerical simulations were carried out and a set of measurements were conducted in an anechoic chamber, where two walls were installed. The experimental setup is shown in Figure 4.

A. Experimental setup

1) Cylindrical loudspeaker array: The prototype of a CLA with radius $a = 0.25\,\text{m}$ and $L = 32$ *HiVi B1S* full-range loudspeakers [43] with a magnesium-aluminium alloy diaphragm of $34\,\text{mm}$ in diameter was located in the anechoic chamber at the position shown in Figure 4. All loudspeakers were calibrated relatively to a reference one.

2) Reflecting walls: Two wooden boards of $3.6\,\text{m} \times 2.4\,\text{m}$ were arranged vertically to create specular wall reflections. To obtain the reflection coefficient of the walls, the loudspeakers' impulse responses were measured both with and without the walls installed. The results revealed that the walls in the anechoic chamber can be assumed almost perfectly rigid, which results in $\xi_1 = \xi_2 = 1$ at all frequencies.

3) Sound field microphone: A *Core Sound TetraMic* [44] was set at the control position to measure the sound field. The microphone was mounted on a stand attached to an electronic turntable and was facing the direction-of-arrival of the target plane wave. This sensor was used to measure the 0th and 1st order terms of the spherical harmonic representation of the reproduced field, also referred to as 1st order Ambisonic B-format signal. The 0th order term, W corresponds to the acoustic pressure, whilst the first order terms $X, Y,$ and Z are proportional to the three components of the particle velocity.

B. Design of Input filters

The 32 loudspeaker filters were computed in the frequency domain with the method described in Section IV. These were
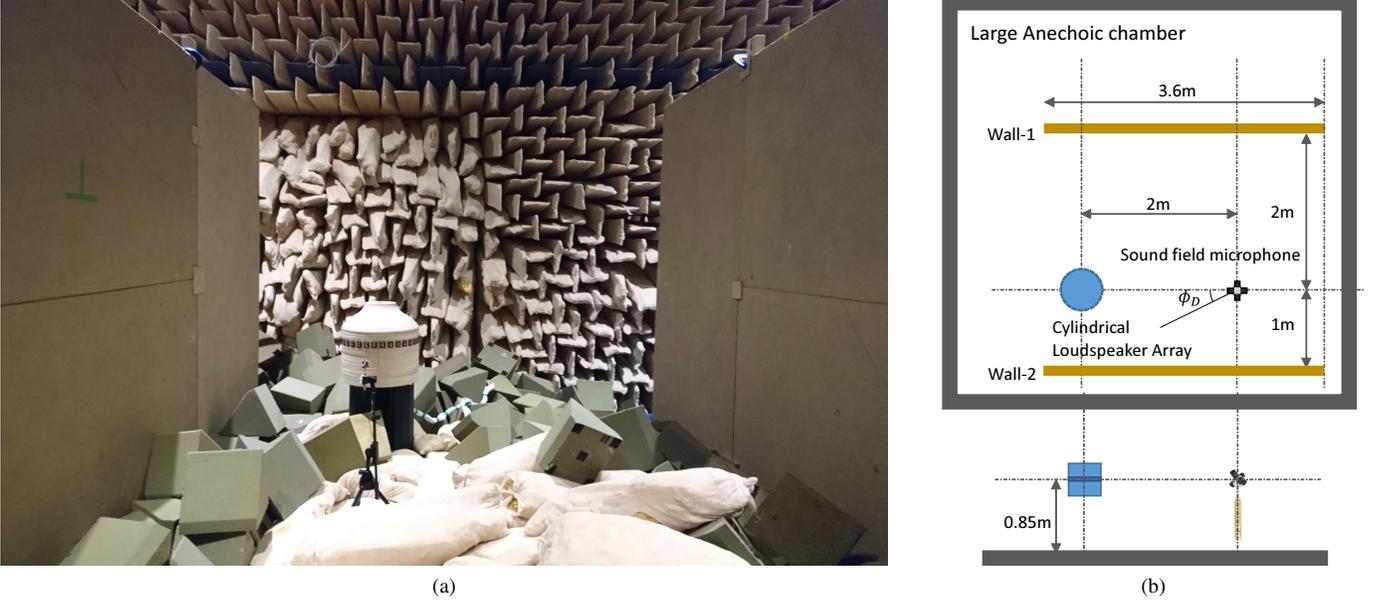
Fig. 4: (a) Experimental setup: a Cylindrical loudspeaker array and a sound field microphone are placed between two walls in an anechoic chamber. (b) Position of the CLA and two walls.

then used to filter a single input test signal (an exponential sinusoidal sweep) and obtain the loudspeaker test signals. The reproduced field was measured and the system impulse responses were obtained by convolving the measured signals with the inverse filter for the specific test signal used, as defined by Farina in [45], [46]. It should be mentioned that the direct inversion of the mode transfer matrix $\mathbf{\Psi}$ in Equation (26) would have resulted in large signal amplification at low frequencies, corresponding to a large amount of energy required to generate the desired sound field while controlling higher orders. Furthermore, this large amount of energy in combination with small errors in the estimation of the CLA and walls positions, of the wall reflection coefficient, or of other parameters that contribute to the transfer functions may have resulted in large differences between the desired sound field and the reproduced one [21]. These effects were mitigated by using the technique of normalized Tikhonov regularization [47], also preventing distortion caused by extreme voice coil and diaphragm excursions and ensuring robustness and stability of the system.

The modal coefficients of the directional sound source were thus obtained as

$$\hat{\mathbf{q}} = \mathbf{\Psi}^H [\mathbf{\Psi}\mathbf{\Psi}^H + \beta\sigma^2\mathbf{I}]^{-1}\mathbf{b}, \qquad (35)$$

where β is the regularisation parameter, σ is the largest singular value of $|\mathbf{\Psi}|$ and $\mathbf{I}$ is the 3×3 identity matrix. Note that vector $\mathbf{b}$ includes only three elements, corresponding to the coefficients of the spherical harmonics Y_0^0, Y_1^{-1}, and Y_1^1. Since the matrix $\mathbf{\Psi}$ is formed from the radial functions and Green's function, its elements depend on the wavenumber, k, and thus on the speed of sound, c. The latter changes with temperature, and therefore the loudspeaker signals must be adjusted accordingly, should the temperature change, e.g. from a lookup table. The ability to account for this makes the system

adaptable to temperature variations, as it can update the loudspeaker signals from their analytical expressions accordingly.

Using Equation (35), different sets of filters for both different target plane wave angles and beam pattern orders $M = 2$ and $M = 7$ were created and their performance compared.

As discussed in Section IV-B, the proposed method relies on the creation of several sound beams, one for each image source. This is the case also if the method is formulated as direct mode matching (refer to Section IV-A), as the two methods were shown to be equivalent. The gains w_l of the three beams as a function of the target plane wave direction are reported in Figure 5. It is shown that when the beams correspond to the direction of either the CLA or of an image source (vertical dotted lines), as seen by the listener, only one beam is generated. When the target angle is between the CLA and an image source, the two beams corresponding to these two directions are mainly active, whereas the third beam has a very low and negative gain. When the target angle is outside the span of the two image sources, all three beams are active. The two image source beams have positive gains, whereas the direct-sound beam has a negative gain, which indicates the occurrence of strong destructive interference between the beams at the listener position.

As an example, Figure 6 shows the magnitude of the Frequency Response Function (FRF) and the Impulse Response (IR) of the filter associated to the 5th loudspeaker, with $M = 2$. The three distinct pulses in Figure 6 (d) are required for generating the three different sound beams, as above. The first and second pulses are generating the sound associated to the wall reflections, whilst the third pulse is for the beam directed towards the listener (direct sound). This pulse is delayed with respect to the other one to account for the path difference between the direct and reflected sound.
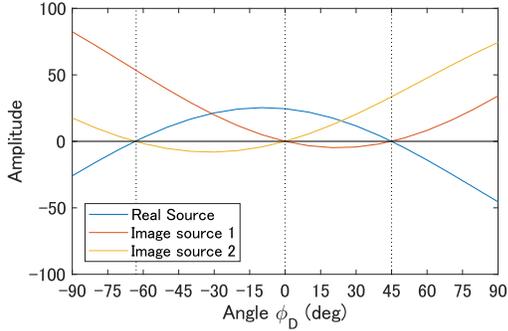
Fig. 5: Beam gains w_l for varying target angle ϕ_D.
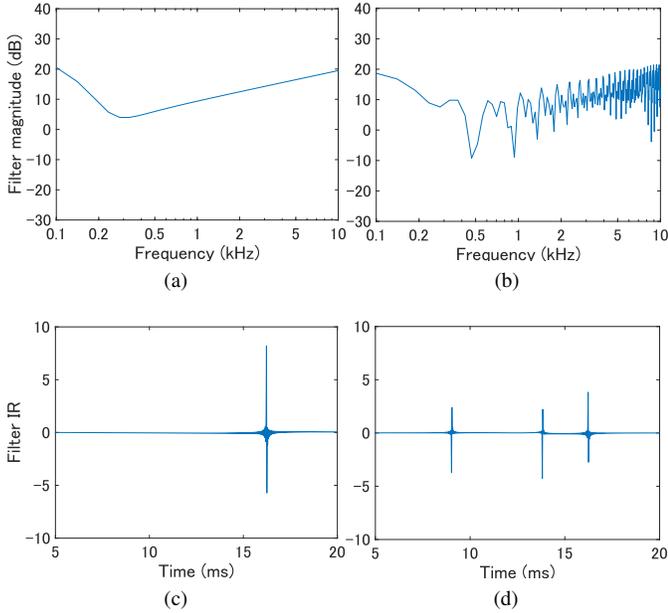
Fig. 6: Frequency response (a,b) and impulse response (c,d) of the fifth loudspeaker's filter with $M = 2$, with (a,c) $\phi_D = 0°$ and (b,d) $\phi_D = 30°$.

C. Beam Pattern

Figure 7 shows both the simulated and the measured beam pattern. This is defined as the sound field generated by the CLA in free field and in the absence of the walls, measured on a 2 m-radius circle in the loudspeaker plane ($z = 0$) and centred on the CLA axis. To measure the beam pattern, the CLA was attached to a turntable and was rotated with steps of 5°, and the radiated sound was recorded with a *Brüel & Kjær* 1/2-inch free-field microphone set at 2 m distance. Each column in Figure 7 corresponds to a different target source direction, namely 0°, 45°, and 60°, respectively, as seen by the listener. The first two rows report the magnitude spectra of the simulated beam patterns for orders $M = 2$ and $M = 7$, respectively. The third row reports the measured beam pattern for $M = 7$, while the last row reports the simulated beam pattern in the time domain, i.e. the impulse response as a function of the radiation angle, for $M = 7$. In all figures, the vertical axis represents the angle $\phi^{(0)}$, the azimuth consistent with the CLA reference system (refer to Figure 2). The listener

is at 180°, while the reflection points on the two walls are at 135° and 243°, respectively. A reflection point is defined as the intersection between a wall and the line joining the listener and the image source corresponding to that wall. The reflection point angles, as seen by the listener, correspond to the image source angles (45° and −63°).

With reference to Figure 7d, it can be observed that the beam pattern shows three distinct frequency bands: a mid-frequency band, between 1.25 kHz and 4.09 kHz (calculation based on [32] and [19]), where the pattern is as desired and clearly features a narrow main lobe and several side lobes; a low frequency region, where the main beam gets progressively broader because of the combined result of the array size becoming small in comparison to the wavelength and of regularisation; a high frequency region, where grating lobes of similar magnitude to that of the main lobe appear as a result of spatial aliasing. Comparison of the results corresponding to $M = 7$ with the beam patterns for $M = 2$ (Figure 7a) indicates that the main lobe of the 2nd order beam is much broader than the 7th order one. Also, for the 2nd order beam, the mid-frequency band described above ranges from 347 Hz to 4.91 kHz and is therefore broader than for the 7th order. All these results are well-known and consistent with the literature (see for example [32], [34] and [19]).

The patterns in the first and second column correspond to the case when the target source direction coincides with either the real source (first column) or an image source (second column). In these cases the simulated beam patterns clearly show that, consistently with Figure 5, a single sound beam is generated in the direction of either the listener (first column) or of the reflection point (second column) and that nulls are steered in the other two directions. The third column corresponds to the case when the target source is out of the angular span defined by the two image sources. In this case three beams can be identified, as expected. Note that the beam directions depend only on the system geometry and not on the target plan wave angle, which affects only the beam amplitude. The simulated beam patterns in the exact direction of either the listener or the reflection points do not vary significantly with frequency. On the contrary, the patterns show significant frequency variations at angles between beams. These oscillations are informally referred to as *comb-filter* effects and are caused by constructive and destructive interference between the lobes of different beams. This phenomenon is much more pronounced in the $M = 2$ case (Figure 7c) than for $M = 7$ (Figure 7f).

The experimental results in Figures 7g-i show very good agreement with simulations in Figures 7d-f. The beam pattern is obviously affected by the frequency response of the loud-speakers. Their low-frequency roll-off is clearly observable below 300 Hz. The narrow beams at the various angles are clearly visible as well as the three frequency bands describe above. Main and side lobes are less regular across frequency than in simulations. This is most likely due to unwanted reflections and diffraction due to the array supporting structure, which are not present in simulations (note that the latter assume a rigid cylindrical baffle of infinite length).

The time-domain beam pattern in Figures 7j-l show some

noteworthy features. The beamforming focusing effect is clearly visible, which causes all the loudspeaker wave fronts to add-up constructively in the directions of the three beams. The time delay between the three different beam signals is also clearly visible. As already explained, this is required to account for the path difference between direct and reflected sounds, so that all three beam signals arrive simultaneously at the listener position. The delay also contributes to the comb-filtering effect caused by the interference between beams that was discussed above. The multiple wave fronts generated by the various loudspeakers are clearly observable. The nulls that can be clearly identified in the mid-frequency region in the frequency-domain plots cannot be easily recognised here. This is probably due to the fact that the beam pattern appearance in the time domain is strongly affected by high frequencies, and spatial aliasing may therefore hide some of the features like the beams' nulls. A pre-ringing phenomenon can be observed at angles distant from the beam steering direction. It is likely that this is the time-domain representation of the low-frequency beam widening effect caused by regularisation. It is in fact known that Tikhonov regularisation may cause pre-ringing artefacts as the β regularisation parameter is generally chosen to be real-valued. Masiero and Vorländer proposed a method to overcome this issue [48], but this was not implemented for this work.

D. Impulse response at the listener's position

This section reports the analysis of the sound field reproduced at the listener's position and in the presence of the two walls, as depicted in Figure 4. Figure 8 reports a short portion of the measured impulse responses as a function of the target angle ϕ_D (vertical axis in the figure), for expansion orders $M = 2$ and $M = 7$. The 0 of the time axis was chosen as the time when the desired plane wave is expected to arrive. It is clearly visible that a compact wave front arrives at that time for all target directions. This is followed by smaller wave fronts, occurring within $1.5\,\mathrm{ms}$ after the main wave front. This corresponds to a path difference of less than $0.5\,\mathrm{m}$, suggesting that these reflections may be caused by scattering off the array supporting structure (consider that the baffle is not an infinitely long cylinder as in the theoretical model) or by some of the non-ideal absorptive wedges arranged on the floor (see Figure 4). Observing again the IR at exactly $0°$, no pre-ringing can be detected due to the fact that only the direct beam is generated. On the other hand, the marked reflection at $2.4\,\mathrm{ms}$ is mainly caused by grating lobes caused by spatial aliasing and, to a lesser extent, by the broad main beam at low frequencies, which are reflected by the closer wall and reach the listener position after the main wave front. This hypothesis is confirmed by the spectral analysis of that specific portion of the IR (not reported in this paper). Also, a closer analysis of these wavelets shows they are constituted by several pulses arriving at different times instead of a single coherent wave front. A similar group of reflections, caused by the farther wall, occurs at $7.2\,\mathrm{ms}$, but is not reported in the figure.

As the target direction progressively moves away from $0°$, other beams, aimed at the reflection points, are generated

(refer to Figure 5). These beams have grating lobes in the listener direction, which arrive before the main wave front. More specifically, the peaks at $-2.4\,\mathrm{ms}$ correspond to the beam associated to the image source at $45°$ "leaking" into the listener direction, whereas the peaks at $-7.2\,\mathrm{ms}$ correspond to the grating lobes of the beam directed toward the other wall, with an image source at $-63°$. Again, the spectral analysis of these portions of the IR confirm that these are caused mainly by spatial aliasing. The peaks at $-4.8\,\mathrm{ms}$ are due to the beams of the $-63°$ image source being reflected by the opposite wall, corresponding to the $45°$ image source. The comparison of Figure 8a and 8b clearly indicates that these unwanted reflections are much more significant for the second order filters.

Figure 9 reports a longer portion of the impulse response than in Figure 8, for a single target angle $\phi_D = 30°$, for both $M = 2$ and $M = 7$. Besides the features described in the previous paragraphs, these plots also show the higher order reflections, not included in Figure 8. It can be clearly seen that these are much more prominent for $M = 2$ than for $M = 7$, consistently with the fact that the 7th order filters allow for a much higher beam pattern directivity, and thus for a smaller amount of acoustic energy being injected into the environment and for a higher direct-to-reverberant sound ratio.

E. Reproduction error and angular error

The accuracy of the reproduced sound field is assessed by means of the sound field reproduction error. For the purpose of this work, this is computed as the squared l_2-norm of the difference between the vectors with the three desired and reproduced spherical harmonic coefficients, normalised by the energy of the target coefficient vector. Mathematically this is defined as

$$E_{\mathrm{rep}} = \frac{||\mathbf{b} - \mathbf{b}_R||^2}{||\mathbf{b}||^2}, \qquad (36)$$

where $\mathbf{b}$ is the vector of target sound field coefficients, as defined in Section IV, and $\mathbf{b}_R$ is the vector of the coefficients measured by the *TetraMic* [44].

In order to assess the directional accuracy of the reproduced sound field, the active intensity and the angular error E_{ang} of the reproduced sound intensity are calculated as follows:

$$\boldsymbol{I}_R = \mathrm{Re}\left[WX^*, WY^*, WZ^*\right], \qquad (37)$$

$$E_{\mathrm{ang}} = \arccos \frac{\boldsymbol{I}_T \cdot \boldsymbol{I}_R}{||\boldsymbol{I}_T||\,||\boldsymbol{I}_R||}, \qquad (38)$$

where W, X, Y, Z are the B-format signals recorded by TetraMic, $\boldsymbol{I}_T$ is the target active sound intensity and $\boldsymbol{I}_R$ is the active intensity of the reproduced sound field at the measurement position. A similar definition of angular error was already used in previous work [49].

The reproduction error and the angular error of the sound field reproduced by the CLA generating 2nd order and 7th order beams are shown in Figure 10. The horizontal axis represents frequency and the vertical axis represents the incident angle ϕ_D of the target plane wave. A $30\,\mathrm{ms}$ Tukey window was applied to the measured impulse responses to pick up only early reflections and discard later reflections from both error
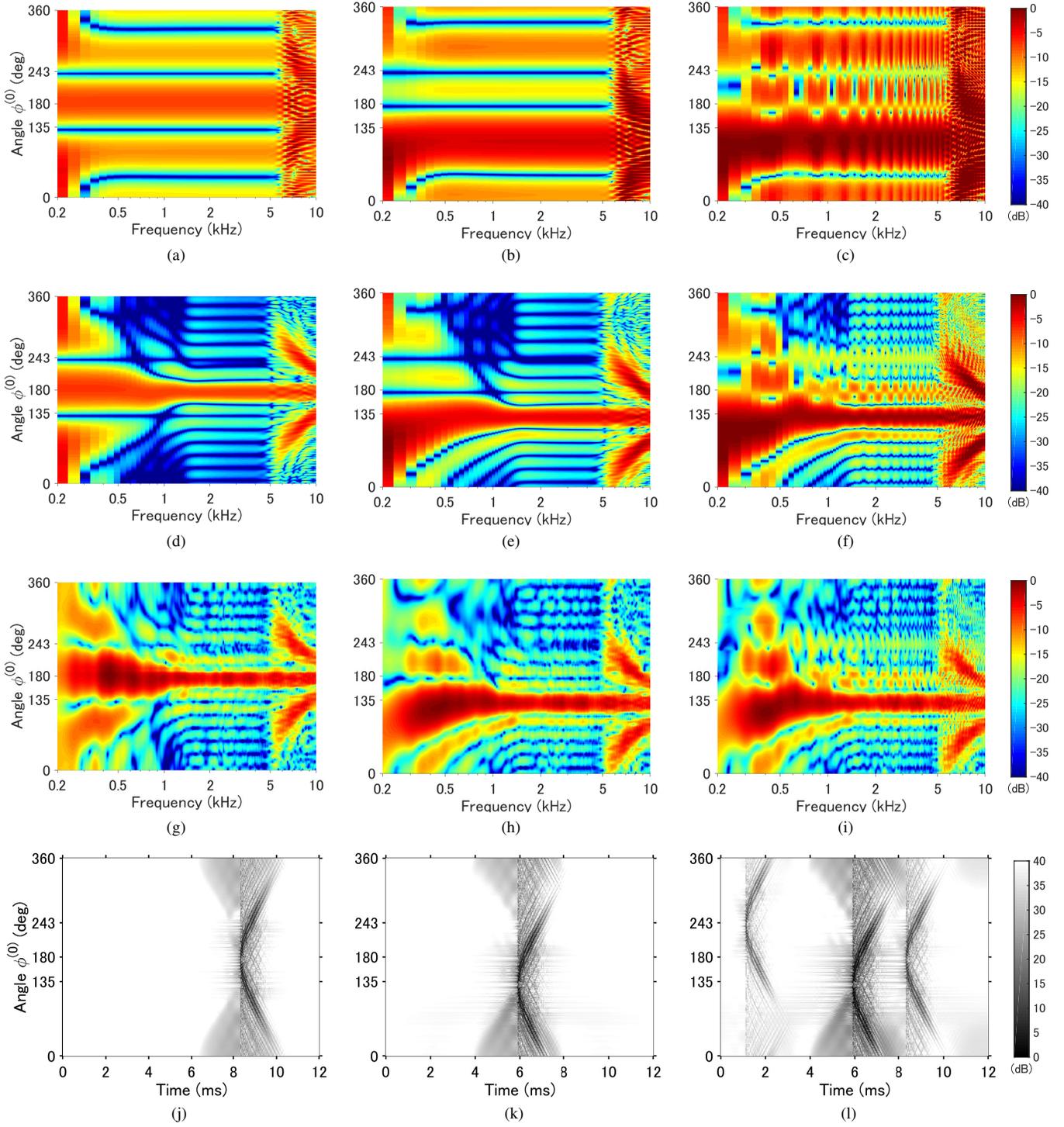
Fig. 7: Beam pattern. (a-f) simulated frequency-domain beam pattern with (a-c) $M = 2$ and (d-f) $M = 7$. (g-i) measured frequency-domain beam pattern with $M = 7$. (j-l) simulated time-domain beam pattern with $M = 7$. The target plane wave directions are: (a,d,g,j) $\phi_D = 0°$, (b,e,h,k) $\phi_D = 45°$, and (c,f,i,l) $\phi_D = 60°$.

(a)

(b)
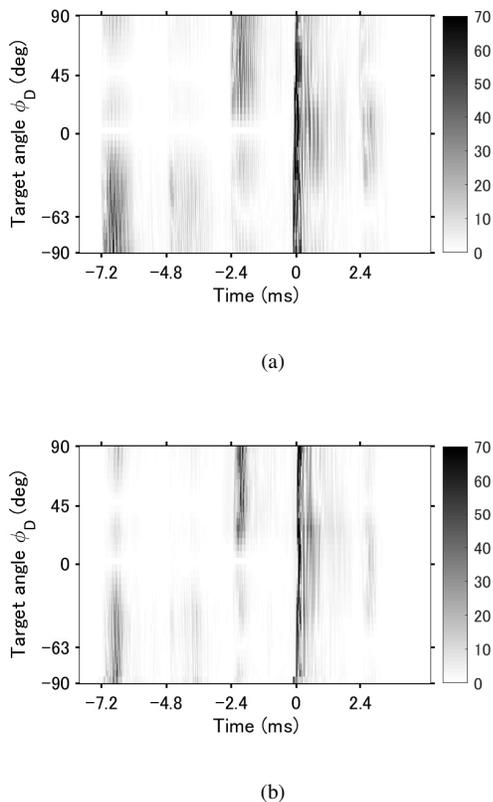
Fig. 8: Impulse response at the control position (a) $M = 2$. (b) $M = 7$.

(a)

(b)
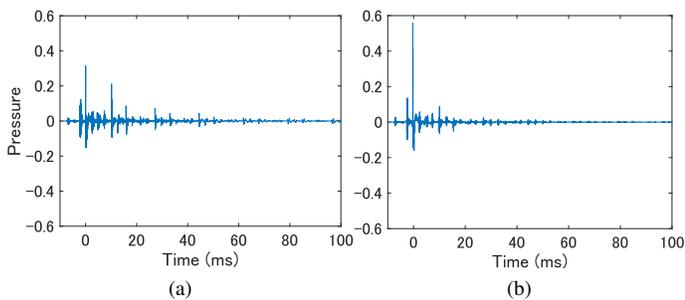
Fig. 9: Impulse response at the control position for $\phi_D = 30°$ (a) $M = 2$. (b) $M = 7$.

calculations. The reason for this choice is that, as mentioned before, no attempt is made to control the reverberant field. Furthermore, the computed error is expected to correlate to human sound localisation cues, especially at high frequencies, and these cues are not affected by late reflections [26].

The reproduction performance depends on the position of the CLA, on the 1st order image sources, on the order of the beams, and on the target angle. The error is smallest in the direction of the CLA (0°) and of the image sources (45° and −63°), where the CLA generates only one sound beam. For directions within the image source span, the error is relatively small. The 7th order filters allow for a lower error at high frequencies, especially between −63° and 0° . For the out-of-span angles ($\phi_D > 45°$ and $\phi_D < −63°$) the error increases
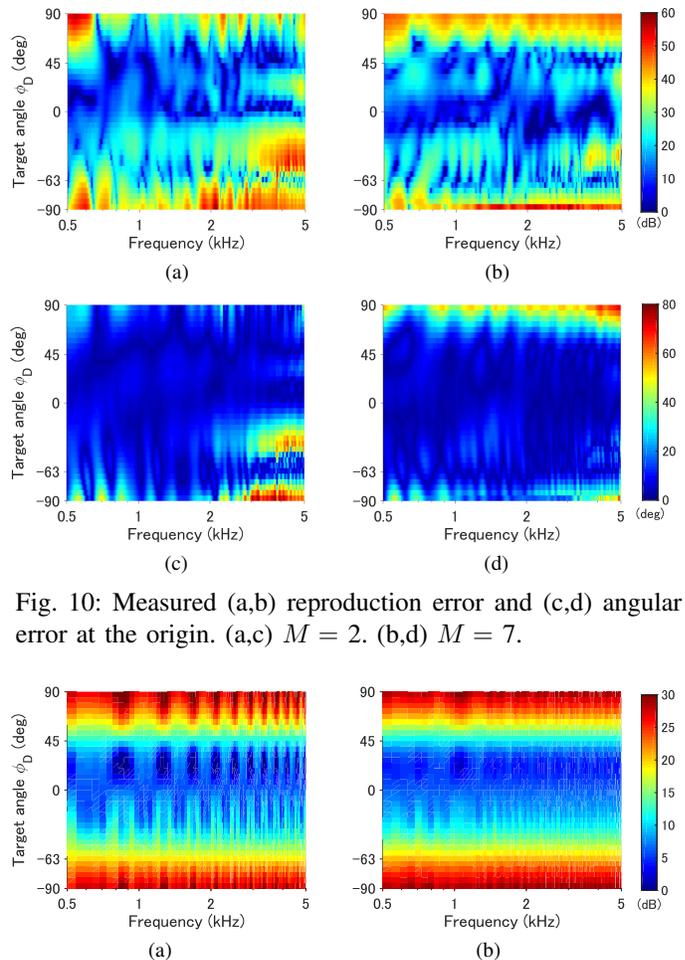
Fig. 10: Measured (a,b) reproduction error and (c,d) angular error at the origin. (a,c) $M = 2$. (b,d) $M = 7$.
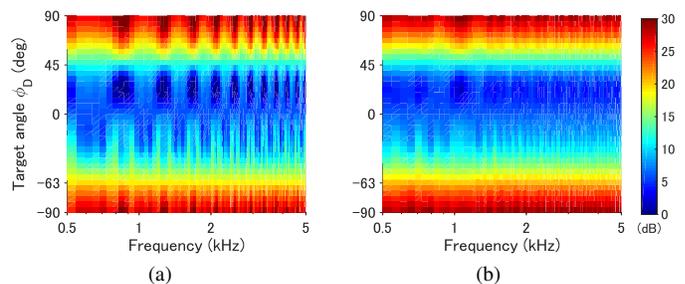
Fig. 11: Array effort. (a) $M = 2$. (b) $M = 7$.

considerably, and the 2nd order filters seem to perform slightly better than the 7th order ones. As explained above, this is the region were significant destructive interference between the three beams occurs (see Figure 5), and it can therefore be expected that the system is more prone to errors such as position inaccuracies, loudspeaker mismatches, changes of temperature and humidity, presence of unwanted reflections, etc. This reasoning is corroborated by the correlation between the measured errors and the array effort, which evaluates the loudspeaker energy required and is related to the robustness of the system. It is defined as,

$$AE = \frac{\sum_{l=1}^{L} |q_l|^2}{|q_{\text{ref}}|^2},\qquad(39)$$

where q_{ref} is the mass acceleration of a single monopole at the centre of the CLA required to produce the same sound pressure at the evaluation point. The AE reported in Figure 11 is a function of both frequency and target plane wave angle and is very large when the latter is out of the image source span. This correlation between the E_{rep}, E_{ang} and the array effort indicates that reproducing out-of-span virtual sources is both inefficient and less robust, and that small errors may lead to large deviations from the target sound field. This is consistent with what happens with the reproduction of out-

of-span image sources with standard stereo and multi-channel audio [50].

The asymmetry of the errors in Figure 10 may be caused by the fact that the wall generating the image source at $-63°$ is farther from the CLA than the wall of the $45°$ image source ($2\,\text{m}$ versus $1\,\text{m}$). While increasing the distance between CLA and wall allows for a wider in-span range for ϕ_D, it also seems to increase the possibility of reproduction inaccuracies in that range.

F. Comparison of filter order

The impact of the filter order on the performance of sound field reproduction is discussed in the following, where the two cases of $M = 2$ and $M = 7$ are considered. As shown in Figure 10, the sound field reproduction using 7th-order directional sources results in a better performance than that with 2nd-order ones when the target plane wave arrives from directions within the 1st-order image source span, especially at high frequencies. On the other hand, the 2nd-order filters seem to perform marginally better in the out-of-span region, consistently with the array effort plot in Figure 11.

The most significant difference between the two filter orders is illustrated by Figures 8 and 9. It can be observed that, as a results of the poorer directivity of the 2nd-order filters, these cause much more significant pre-ringing artefacts (see explanation above) and late reflections. The poorer direct-to-reverberant ratio makes the system with 2nd-order filters more vulnerable to room reverberation. The presence of strong pre-ringing artefacts is potentially concerning. The work by Pastore & Braasch has shown that the precedence effect can be overcome with a shift of localisation towards the source of the lagging pulse for long-duration stimuli and for lagging pulses exceeding the leading pulses by $10\,\text{dB}$ or more in level [51]. Note that while their work only studied level differences up to $10\,\text{dB}$ and lag times of up to $5\,\text{ms}$, their data still indicates a trend that a yet larger difference would lead to a localisation towards the source of the louder, lagging pulse. At the same time, increased lag time works against this effect, so it remains desirable to keep the lag time as short as possible.

This means that using 7th-order filters is likely to improve localisation performance by reducing the possibility of the listener perceiving the sound as coming directly from the array. As pre-ringing artefacts may be stronger in certain frequency ranges, it is likely that different parts of the source spectrum may be associated to different angular directions. This may then reduce the spatial accuracy and width of the virtual source, effecting a source perceived as smeared across space. These theoretical considerations would of course need to be validated by subjective listener experiments, which are, however, not included in this work.

G. Elevation

The focus of the experimental analysis has been on the case of $\theta_D = \theta_l = \pi/2$, $\forall l$, which means that all loudspeakers lay in the same plane as the listener (i.e. the center of the reference system) and that the target plane wave is intended to also travel in that plane. While this choice was made to simplify the analysis of the system performance, it must be clarified that the proposed method can account for situations when the listener is not located in the same plane as the loudspeakers and when the target plane wave has a different elevation angle.

Numerical simulations as well as intuitive considerations suggest that the reproduction accuracy is higher when the wave vector of the target plane wave lies in the same plane as all sources and the listener. This plane always exists when two vertical parallel walls are considered, as the two image sources and the center of the CLA lie on the same straight line. When the two walls are vertical but not parallel, as in the example of Figure 2, the reproduction accuracy is highest when the target wave vector lies in the plane containing the listener position, the CLA, and one of the image sources, and points in the direction between the CLA and the image source. As the target plane wave direction moves away from this optimal range, the accuracy of the reproduction progressively reduces.

VI. Conclusion

An approach for sound field reproduction with a Cylindrical Loudspeaker Array (CLA) in a reverberant environment has been presented in this paper. Two techniques, the sound-field-focused direct mode matching and the more intuitive beam steering, have been introduced and shown to be mathematically equivalent. These methods enable the reproduction of a target sound field at a listener position by using little a-priori knowledge of the room's acoustic conditions, namely the distance of the CLA to the surrounding walls and their absorption coefficients, as well as their angles to the reference coordinate system.

The method relies on a simplified acoustic model that considers the radiation from a CLA and takes into account early reflection of nearby walls. The overall pressure field at the listener position can then be approximated in analytical form. This has the advantage that the loudspeaker acoustic transfer functions do not need to be measured at the listening position. Also, to increase the system's robustness, no direct attempt is made to control late reverberation.

To evaluate the performance of this approach, simulations and experiments with the CLA and two walls installed in an anechoic chamber were conducted. The impact of the modal order of the CLA, corresponding to maximum directivity of the system, was investigated. Good results and high robustness (i.e. low array effort) were obtained when reproducing a plane wave coming from a direction that, from the listener's perspective, lies between the two 1st-order image sources. Outside these directions, performance is suffering from inaccuracies in the position of the CLA and essential model parameters, e.g. the speed of sound. Adding more CLAs around the control point and measuring the temperature to adjust the internal model parameters can then improve performance and robustness.

An extension of this approach to 3D sound field reproduction and subjective tests to determine the minimum CLA directivity required for perceptually robust localisation within the in-span range of directions remain subjects of future work. The behavior of the system and its robustness when

walls absorption depends significantly on frequency or when reflections are also diffuse remain aspects to be investigated, as well as the possibility to compensate for this departure from ideal conditions by means of a simple calibration process.

APPENDIX

In [35], Williams presented a generic formula for the SPA of integrals with the form

$$I(R) = \int_{-\infty}^{\infty} f(k_z) e^{iRg(k_z)} dk_z \tag{40}$$

that is given by

$$I(R) \approx f(k_z) e^{iRg(k_z)} e^{-i\frac{\pi}{4}} \sqrt{\frac{2\pi}{R|g''(k_z)|}}, g''(k_z) < 0 \tag{41}$$

where k_z denotes the stationary phase point that satisfies

$$\left.\frac{dg(k_z)}{dk_z}\right|_{k_z = k_{z0}} = 0 \tag{42}$$

and $g''(k_z)$ is the second derivative of g with respect to k_z. If $r \to \infty$, then the integral in Equation (7) can be rewritten in the form in Equation (40) by replacing the Hankel function by its large argument asymptote [35]

$$H_\mu(x) \approx \sqrt{\frac{2}{\pi x}} e^{i(x - \mu\pi/2 - \pi/4)}. \tag{43}$$

REFERENCES

[1] A. Berkhout, D. D. Vries, and P. Vogel, "Acoustic control by wave field synthesis," *Journal of the Acoustical Society of America*, vol. 93, no. 5, p. 2764–2778, 1993.

[2] J. Daniel, S. Moreau, and R. Nicol, "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," in *Audio Engineering Society Convention 114*. Audio Engineering Society, 2003.

[3] J. Ahrens and S. Spors, "An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 988–999, 2008.

[4] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004–1025, 2005.

[5] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 2992–3000, 1993.

[6] P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Inverse filter design and equalization zones in multichannel sound reproduction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, 1995.

[7] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multi-channel systems using regularisation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, 1998.

[8] O. Kirkeby and P. A. Nelson, "Digital filter designfor inversion problemsin sound reproduction," *Journal of the Audio Engineering Society*, vol. 47, no. 7/8, p. 583–595, 1999.

[9] M. Kolundzija, C. Faller, and M. Vetterli, "Reproducing sound fields using mimo acoustic channel inversion," *Journal of the Audio Engineering Society*, vol. 59, no. 10, p. 721–734, 2011.

[10] L. D. Fielder, "Practical limits for room equalization," in *Audio Engineering Society Convention 111*. Audio Engineering Society, 2001.

[11] S. Spors, H. Buchner, and R. Rabenstein, "A novel approach to active listening room compensation for wave field synthesis using wavedomain adaptive filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, 2004.

[12] H. Buchner, S. Spors, and R. Rabenstein, "Efficient active listening room compensation for wave field synthesis," in *Audio Engineering Society Convention 116*. Audio Engineering Society, 2004.

[13] S. Petrausch, S. Spors, and R. Rabenstein, "Simulation and visualization of room compensation for wave field synthesis with the functional transformation method," in *Audio Engineering Society Convention 119*. Audio Engineering Society, 2005.

[14] M. A. Poletti, T. Betlehem, and T. D. Abhayapala, "Higher-order loudspeakers and active compensation for improved 2d sound field reproduction in rooms," *Journal of the Audio Engineering Society*, vol. 63, no. 1/2, p. 31–45, 2015.

[15] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, 2015.

[16] P. Lecomte, P.-A. Gauthie, C. Langrenne, A. Berry, and A. Garcia, "Cancellation of room reflections over an extended area using ambisonics," *Journal of the Acoustical Society of America*, vol. 143, no. 2, pp. 811–828, 2018.

[17] M. A. Poletti, F. M. Fazi, and P. Nelson, "Sound-field reproduction systems using fixed-directivity loudspeakers," *Journal of the Acoustical Society of America*, vol. 127, no. 6, pp. 3590–3601, 2010.

[18] ——, "Sound-field reproduction systems using variable-directivity loudspeakers," *Journal of the Acoustical Society of America*, vol. 129, no. 3, pp. 1429–1438, 2011.

[19] M. A. Poletti and T. Betlehem, "Design of a prototype variable directivity loudspeaker for improved surround sound reproduction in rooms," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control-Engineering and Perception.* Audio Engineering Society, 2013.

[20] T. Betlehem, C. Anderson, and M. A. Poletti, "A directional loudspeaker array for surround sound in reverberant rooms," in *Proceedings of 20th International Congress on Acoustics, ICA 2010.* International commission for acoustics, 2010.

[21] T. Betlehem and M. A. Poletti, "Two dimensional sound field reproduction using higher order sources to exploit room reflections," *Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 1820–1833, 2014.

[22] N. Deppisch, T. Meyer-Kahlen, F. Zotter, and M. Frank, "Surround with depth on first-order beam-controlling loudspeakers," in *Audio Engineering Society Convention 144*. Audio Engineering Society, 2018.

[23] A. Canclini, D. Marković, F. Antonacci, A. Sarti, and S. Tubaro, "A room-compensated virtual surround system exploiting early reflections in a reverberant room," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE, 2012.

[24] D. Marković, K. Kowalczyk, F. Antonacci, C. Hofmann, A. Sarti, and W. Kellermann, "Estimation of acoustic reflection coefficients through pseudospectrum matching," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 1, pp. 125–137, Jan. 2014.

[25] V. Erbes and S. Spors, "Localisation properties of wave field synthesis in a listening room," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, no. 1, pp. 1016–1024, Mar. 2020.

[26] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization.* The MIT Press, 1996.

[27] S. Takumai, "Loudspeaker array device and method for setting sound beam of loudspeaker array device," Europe Patent 1 760 920 A1, 03 07, 2007. [Online]. Available: https://patents.google.com/patent/EP1760920A1/en

[28] H. Teutsch and W. Kellermann, "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," *Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2724–2436, 2006.

[29] M. Møller, M. Olsen, F. T. Agerkvist, J. Dyreby, and G. K. Munch, "Circular loudspeaker array with controllable directivity," in *Audio Engineering Society Convention 128*. Audio Engineering Society, 2010.

[30] M. Kolundzija, C. Faller, and M. Vetterli, "Baffled circular loudspeaker array with broadband high directivity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, 2010.

[31] ——, "Design of a compact cylindrical loudspeaker array for spatial sound reproduction," in *Audio Engineering Society Convention 130*. Audio Engineering Society, 2011.

[32] F. M. Fazi, M. Shin, F. Olivieri, and S. Fontana, "Low frequency performance of circular loudspeaker arrays," in *Audio Engineering Society Convention 138*. Audio Engineering Society, 2015.

[33] F. M. Hoffmann, F. M. Fazi, E. G. Williams, and S. Fontana, "A general radiation model for sound fields and nearfield acoustical holography in wedge propagation spaces," *Journal of the Acoustical Society of America*, vol. 142, no. 3, pp. 1249–1260, 2017.

[34] F. M. Hoffmann, E. G. Williams, F. M. Fazi, and S. Fontana, "An analytical model for wedge-shaped acoustic arrays," *Journal of Sound and Vibration*, vol. 439, pp. 56–76, 2019.

[35] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. New York: Academic press, 1999.

[36] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[37] J. Borish, "Extension of the image model to arbitrary polyhedra," *Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1827–1836, 1984.

[38] E. A. Lehmann and A. M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269–277, 2008.

[39] H. Kuttruff, *Room Acoustics*. CRC Press, 2017.

[40] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer, 2007.

[41] T. Betlehem and M. A. Poletti, "Sound field of a directional source in a reverberant room," in *21st Biennial ASNZ Conference*. New Zealand Acoustics, 2012.

[42] F. M. Fazi, M. Shin, F. Olivieri, S. Fontana, and L. Yue, "Comparison of pressure-matching and mode-matching beamforming methods for circular loudspeaker arrays," in *Audio Engineering Society Convention 137*. Audio Engineering Society, 2014.

[43] HiVi B1S Full-range Driver. [Online]. Available: https://swanspeakers.com/product/b1s-full-range-driver/

[44] Core Sound TetraMic. [Online]. Available: https://www.core-sound.com/TetraMic/1.php

[45] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

[46] ——, "Advancements in impulse response measurements by sine sweeps," in *Audio Engineering Society Convention 122*. Audio Engineering Society, 2007.

[47] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion (Monographs on Mathematical Modeling and Computation, Series Number 4)*. Society for Industrial and Applied Mathematics, 1987.

[48] B. Masiero and M. Vorländer, "A framework for the calculation of dynamic crosstalk cancellation filters," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 9, pp. 1345–1354, Sep. 2014.

[49] M. Shin, P. A. Nelson, F. M. Fazi, and Seo.J, "Velocity controlled sound field reproduction by non-uniformly spaced loudspeakers," *Journal of Sound and Vibration*, vol. 370, pp. 444–464, 2016.

[50] J. Hollebon and F. M. Fazi, "Generalised low frequency 3d audio reproduction over loudspeakers," in *Audio Engineering Society Convention 148*. Audio Engineering Society, 2020.

[51] M. T. Pastore and J. Braasch, "The precedence effect with increased lag level," *The Journal of the Acoustical Society of America*, vol. 138, no. 4, pp. 2079–2089, Oct. 2015. [Online]. Available: https://doi.org/10.1121/1.4929940