FAST SOUND FIELD REPRODUCTION IN BOX-SHAPED ROOMS: RIGID WALLS CASE

Jorge Martinez[1], *Geert Leus*[1] *and W. Bastiaan Kleijn*[1,2]

[1] Circuits and Systems, Delft University of Technology, The Netherlands
[2]School of Engineering and Computer Science, Victoria University of Wellington, New Zealand

ABSTRACT

In this paper, an approach to sound field reproduction in reverberant rooms is presented. We focus on box-shaped rooms with fully reflective (rigid) walls. We propose a scheme based on mode-matching in the spatio-temporal Fourier domain combined with a simple least-squares approach to derive the loudspeaker weights that render the target sound-field. By taking advantage of the fast Fourier transform (FFT), the method leads to a fast way to compute the loudspeaker weights. We address the reconstruction of basic sound fields (room-modes) using linear loudspeaker array configurations. This is important, as more complex sound fields can be decomposed into a set of weighted room-modes. Our simulations show that accurate and efficient reconstruction of room-modes in a reverberant environment with perfectly reflective walls is possible.

Index Terms— sound field synthesis, Fourier domain, mode-matching, box-shaped room, room-modes.

1. INTRODUCTION

The main application of sound field reproduction is to recreate sound fields using an array of loudspeakers to give the impression to the listeners of a predefined acoustic scene. The problem is to find the loudspeaker driving functions such that their emitted sound fields combine to approximate the desired sound field. Approaches to achieve this are traced back to the late 1980's with a method by Berkhout called "Wave Field Synthesis" (WFS) [1], and later methods called "Higher Order Ambisonics" (HOA) (e.g., [2]). In [3], a comprehensive treatise on analytic methods for sound field reproduction is given. These methods are normally derived assuming a *free-field* setup. In a room this assumption is invalid.

In a reverberant environment, the derivation of the loudspeaker weights and accurate reconstruction of the desired sound field poses a challenging problem. One general approach to derive the loudspeaker weights in a reverberant environment consists of a sound-pressure-matching approach based on the inversion or equalization of the room transfer function from each of the loudspeaker positions to a dense set of points in the listening area. This easily leads to ill-conditioned solutions given the non-minimum-phase properties of the room transfer function, and to fast divergence from the desired solution in uncontrolled zones [4]. One of the first methods to achieve sound field synthesis in a controlled reverberant environment is given in [5]. More recent papers combine models of wave-propagation with feedback obtained by placing microphones in the desired reconstruction zone to correct for any discrepancies between the reconstructed and desired sound field [6]. Generally, the complexity of these methods increases drastically with the target temporal frequency. More importantly, the synthesis of *room-modes* – the basis functions into which any sound field in a room can be decomposed – has not been widely treated in the literature [3].

In this paper we address the reproduction of virtual sound fields in a room using an array of loudspeakers. We use the model introduced in [7] where reverberation is characterized by discretization in the spatio-temporal Fourier domain. We then apply mode-matching in the discrete Fourier domain to derive the loudspeaker weights and combine the obtained analytic solution with a simple least-squares approach. For simplicity of presentation, 2D (height invariant) sound fields are considered. In Sec. 2, we first give a brief introduction to the reverberation model in [7] as it is used in this work. We then use a similar approach as in [8]: Given a desired virtual sound field, the aperture function (i.e. the driving function of a continuous loudspeaker) that would render this sound field is calculated. In Sec. 3, the continuous aperture function is sampled in space, which has the interpretation of replacing the continuous loudspeaker with an array of point-sources. Given an array of loudspeakers arranged in a given geometry we derive the least-squares approximation that best fits the analytic sampled aperture. The computation of the loudspeaker weights is performed efficiently using the fast Fourier transform (FFT). Simulation results are given in Sec. 4.

2. SOUND FIELD REPRODUCTION IN BOX-SHAPED ROOMS

The domain of interest is a box-shaped room with dimensions L_x, L_y and L_z, where one corner of the room is positioned at the origin of the coordinate system and the room walls are arranged perpendicular to each of the coordinates. The room walls are assumed to be rigid (i.e., fully reflective). The analysis is given for 2D height-invariant sound fields, that is, the sound field in the z direction is assumed constant. A 2D point-source is therefore seen as a line-source extending in the z

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

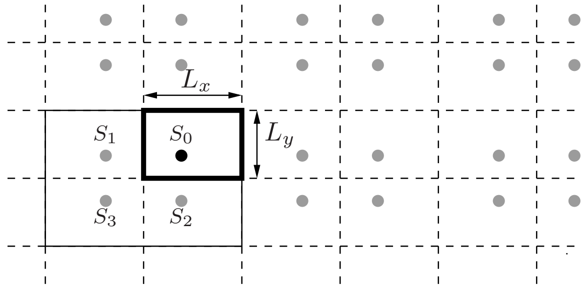


Fig. 1: A source (black circle) in a room. The room space is enclosed by a bold line. Sound reflections are simulated as virtual copies (gray circles) of the source.

direction from 0 to L_z. Fixing the z coordinate to any height z_0, a position in space is represented by $\mathbf{x}=[x,y]^\mathsf{T}$, were superscript $^\mathsf{T}$ indicates vector or matrix transposition. Likewise, a source position is given by $\mathbf{x}_s=[x_s,y_s]^\mathsf{T}$.

Reverberation in the room is modeled by the creation of free-field virtual image sources outside the room representing the reflections due to the walls, an approach known as the mirror image source model (MISM) [9]. In the case of a box-shaped room with fully reflective walls, any sound field can be modeled as spatially periodic over a lattice when extended outside the boundaries of the room [7]. Let the periodicity lattice be given by Λ, with generator matrix $\mathbf{\Lambda}=\mathrm{diag}(2L_x,2L_y)$ [7]. Note that the period is given by $\mathrm{V}_\Lambda(\mathbf{0})$, the Voronoi region at the origin of lattice Λ defined as $V_\Lambda(\mathbf{0})=\{\mathbf{x}\in\mathbb{R}^3:\|\mathbf{x}\|\le\|\mathbf{x}-\mathbf{\Lambda n}\|,\ \forall\mathbf{n}\in\mathbb{Z}^3\}$. In this case $\mathrm{V}_\Lambda(\mathbf{0})=(-L_x,L_x)\times(-L_y,L_y)$, where $\times$ denotes the Cartesian product. Moreover, inside $\mathrm{V}_\Lambda(\mathbf{0})$, the sound field is mathematically modeled as spatially symmetric. Therefore, if $P(\mathbf{x},\omega)$ denotes the steady-state sound field in a room then for $\mathbf{x}_s\in V_\Lambda(\mathbf{0})$

$$P(\mathbf{x},\omega)=P(-\mathbf{x},\omega)=P(-x,y,\omega)=P(x,-y,\omega), \quad (1)$$

where ω denotes temporal frequency.

A depiction of the scenario is given in Fig. 1. The actual room is indicated by a solid bold line, whereas the Voronoi region is the area enclosed by a solid line. A sound field inside the room is generated by a source S_0. The reflections of the sound field on the walls are modeled by virtual copies of the real source (depicted by gray filled circles). The virtual sources denoted by S_1 and S_2 correspond to first order reflections, and S_3 corresponds to a second order reflection. Higher order reflections on the walls are modeled by periodic repetitions of the set of sources inside the Voronoi region.

Let us now denote the desired (virtual) sound field that is to be reproduced in the room by $P^{(d)}(\mathbf{x},\omega)$. Because of spatial periodicity, the desired sound field can be expanded into a spatial Fourier series as follows,

$$P^{(d)}(\mathbf{x},\omega)=\frac{1}{|\mathbf{\Lambda}|}\sum_{\mathbf{q}\in\mathbb{Z}^2}\hat{P}^{(d)}_{\mathrm{V}_\Lambda}(\mathbf{\Phi q},\omega)\mathrm{e}^{j\mathbf{q}^\mathsf{T}\mathbf{\Phi}^\mathsf{T}\mathbf{x}}, \quad (2)$$

where $\mathbf{q}=[q_x,q_y]^\mathsf{T}$ is an integer vector, $|\mathbf{\Lambda}|$ is the absolute value of the determinant of $\mathbf{\Lambda}$, $\mathbf{\Phi}$ is the generator matrix $\mathbf{\Phi}=2\pi\mathbf{\Lambda}^{-\mathsf{T}}$ of the (scaled) reciprocal lattice Φ that defines the spectral sampling points. The superscript $^{-\mathsf{T}}$ denotes the inverse transposed. The Fourier coefficients $\hat{P}^{(d)}_{\mathrm{V}_\Lambda}(\mathbf{\Phi q},\omega)$, are samples of the continuous spatio-temporal spectrum $\hat{P}^{(d)}_{\mathrm{V}_\Lambda}(\boldsymbol{\phi},\omega)$ [7]. Function $\hat{P}^{(d)}_{\mathrm{V}_\Lambda}(\boldsymbol{\phi},\omega)$ is the spatial Fourier transform of $P^{(d)}_{\mathrm{V}_\Lambda}(\mathbf{x},\omega)$, which denotes the desired sound field generated in free-field by the sources in the Voronoi region, and $\boldsymbol{\phi}=[\phi_x,\phi_y]^\mathsf{T}$ is the spatial frequency vector.

The actual sound field in the room (indicated by an $^{(a)}$ superscript) generated by a continuous source emitting a signal $s(\mathbf{x}_s,\omega)$, is completely characterized by the Green's function $G(\mathbf{x},\mathbf{x}_s,\omega)$ from the source position to any point in the room,

$$P^{(a)}(\mathbf{x},\omega)=\int_{V_\Lambda(\mathbf{0})}s(\mathbf{x}_s,\omega)G(\mathbf{x},\mathbf{x}_s,\omega)\mathrm{d}\mathbf{x}_s. \quad (3)$$

Further any valid source function $s(\mathbf{x}_s,\omega)$ has to fulfill the symmetry conditions (1). Therefore, although the integral is evaluated over the entire Voronoi region, the source function only has degrees of freedom inside the room. This function can be seen as the *aperture function* of a continuous loudspeaker inside the room that gets reflected on the other three zones of the Voronoi region (see Fig. 1). Thus, for $\mathbf{x}_s\in V_\Lambda(\mathbf{0})$ we have $s(\mathbf{x}_s,\omega)=s(-\mathbf{x}_s,\omega)=s(-x_s,y_s,\omega)=s(x_s,-y_s,\omega)$.

Given the model presented in [7], the Green's function of the room can be obtained using a sampling scheme on the spatio-temporal Fourier transform of the free-field room impulse response (RIR). In our 2D scenario it takes the form $\hat{G}_{\mathrm{ff}}(\boldsymbol{\phi},\mathbf{x}_s,\omega)=\mathrm{e}^{-j\boldsymbol{\phi}^\mathsf{T}\mathbf{x}_s}(2\pi(\|\boldsymbol{\phi}\|^2-(\omega/c)^2))^{-1}$ [10], where c denotes the velocity of sound propagation. The Green's function of the room can be expanded as a spatial Fourier series using spectral samples of free-field Green's function,

$$G(\mathbf{x},\mathbf{x}_s,\omega)=\frac{1}{|\mathbf{\Lambda}|}\sum_{\mathbf{q}\in\mathbb{Z}^2}\hat{G}_{\mathrm{ff}}(\mathbf{\Phi q},\mathbf{x}_s,\omega)\mathrm{e}^{j\mathbf{q}^\mathsf{T}\mathbf{\Phi}^\mathsf{T}\mathbf{x}}. \quad (4)$$

On the other hand, the source aperture function $s(\mathbf{x}_s,\omega)$ is periodic in space, we can then expand it as a Fourier series

$$s(\mathbf{x}_s,\omega)=\frac{1}{|\mathbf{\Lambda}|}\sum_{\mathbf{q}\in\mathbb{Z}^2}\beta_\mathbf{q}(\omega)\mathrm{e}^{j\mathbf{q}^\mathsf{T}\mathbf{\Phi}^\mathsf{T}\mathbf{x}_s}, \quad (5)$$

with $\beta_\mathbf{q}(\omega)$, the spatial coefficient functions of the aperture.

To derive the unknown aperture coefficients $\beta_\mathbf{q}(\omega)$ from the known coefficients of the desired sound field $\hat{P}^{(d)}_{\mathrm{V}_\Lambda}(\mathbf{\Phi q},\omega)$ in (2), we substitute the series expansions for the RIR (4) and the source (5) into the actual sound field expression (3) i.e.,

$$\begin{aligned}P^{(a)}(\mathbf{x},\omega)&=\int_{V_\Lambda(\mathbf{0})}s(\mathbf{x}_s,\omega)G(\mathbf{x},\mathbf{x}_s,\omega)\mathrm{d}\mathbf{x}_s\\&=\frac{1}{2\pi|\mathbf{\Lambda}|^2}\sum_{\mathbf{q}'\in\mathbb{Z}^2}\sum_{\mathbf{q}\in\mathbb{Z}^2}\beta_{\mathbf{q}'}(\omega)\frac{\mathrm{e}^{j\mathbf{q}^\mathsf{T}\mathbf{\Phi}^\mathsf{T}\mathbf{x}}}{\|\mathbf{\Phi q}\|^2-(\omega/c)^2}\\&\qquad\times\int_{V_\Lambda(\mathbf{0})}\mathrm{e}^{j\mathbf{q}'^\mathsf{T}\mathbf{\Phi}^\mathsf{T}\mathbf{x}_s}\mathrm{e}^{-j\mathbf{q}^\mathsf{T}\mathbf{\Phi}^\mathsf{T}\mathbf{x}_s}\mathrm{d}\mathbf{x}_s. \quad (6)\end{aligned}$$

The integral in (6) is equal to $|\mathbf{\Lambda}|$ if $\mathbf{q}' = \mathbf{q}$ and 0 otherwise. Then we obtain

$$P^{(a)}(\mathbf{x}, \omega) = \frac{1}{|\mathbf{\Lambda}|} \sum_{\mathbf{q} \in \mathbb{Z}^2} \frac{\beta_{\mathbf{q}}(\omega) \mathrm{e}^{j\mathbf{q}^\mathsf{T} \mathbf{\Phi}^\mathsf{T} \mathbf{x}}}{2\pi \left(\|\mathbf{\Phi q}\|^2 - (\omega/c)^2\right)}. \quad (7)$$

Equating (2) and (7) term by term we get the aperture coefficients in terms of the desired sound field coefficients, i.e.,

$$\frac{\beta_{\mathbf{q}}(\omega)}{2\pi(\|\mathbf{\Phi q}\|^2 - (\omega/c)^2)} = \hat{P}^{(d)}_{\mathrm{V}_\Lambda}(\mathbf{\Phi q}, \omega)$$

$$\Rightarrow \quad \beta_{\mathbf{q}}(\omega) = 2\pi \hat{P}^{(d)}_{\mathrm{V}_\Lambda}(\mathbf{\Phi q}, \omega) \left(\|\mathbf{\Phi q}\|^2 - (\omega/c)^2\right). \quad (8)$$

Substituting (8) into (5), the aperture function is given by Fourier synthesis of the obtained coefficients,

$$s(\mathbf{x}_s, \omega) = \frac{2\pi}{|\mathbf{\Lambda}|} \sum_{\mathbf{q} \in \mathbb{Z}^2} \hat{P}^{(d)}_{\mathrm{V}_\Lambda}(\mathbf{\Phi q}, \omega) \left(\|\mathbf{\Phi q}\|^2 - (\frac{\omega}{c})^2\right) \mathrm{e}^{j\mathbf{q}^\mathsf{T} \mathbf{\Phi}^\mathsf{T} \mathbf{x}_s}. \quad (9)$$

This equation defines the signal that a (continuous) loudspeaker in the room must output in order to generate the desired sound field. Note that the source function $s(\mathbf{x}_s, \omega)$ can take values at any point in the space of the room. This is undesirable as we want to place loudspeakers only at convenient locations.

Of interest is, for example, a continuous aperture function defined on a linear spatial manifold. Let a continuous linear loudspeaker be positioned parallel to the y direction at a fixed x_0. In this case the aperture function of the linear loudspeaker, say $\bar{s}(\mathbf{x}_s, \omega)$, inside the room (for $\mathbf{x}_s \in (0, L_x) \times (0, L_y)$) takes the form, $\bar{s}(\mathbf{x}_s, \omega) = \bar{s}(x_s, y_s, \omega) = \delta(x_s - x_0)\bar{s}(y_s, \omega)$ where δ is Dirac's delta function. The source function in the Voronoi region is thus given by two linear loudspeaker distributions extending in the y direction from $y = -L_y$ up to L_y, positioned at x_0 and $-x_0$ respectively. Then for $\mathbf{x}_s \in V_\Lambda(\mathbf{0})$ we have, $\bar{s}(\mathbf{x}_s, \omega) = \big(\delta(x_s - x_0) + \delta(x_s + x_0)\big)\bar{s}(y_s, \omega)$, or from (5) in the spatio-temporal Fourier domain

$$\bar{\beta}_{\mathbf{q}}(\omega) = 2\cos(q_x(\pi/L_x)x_0)\bar{\beta}^{(y)}_{q_y}(\omega), \quad (10)$$

since $\mathbf{\Phi} = \mathrm{diag}(\pi/L_x, \pi/L_y)$, where $\bar{\beta}^{(y)}_{q_y}(\omega)$ are the spatial Fourier coefficients in the y dimension. Our interest is then to compute the coefficients $\bar{\beta}_{\mathbf{q}}(\omega)$ of the linear source that "best" approximate the coefficients $\beta_{\mathbf{q}}(\omega)$ of the loudspeaker function that can take values anywhere in the room. We can pose the approximation problem at this point. It is however advantageous to first analyze the effect of spatial sampling on these continuous loudspeaker functions (interpreted as placing loudspeakers at those sample positions in space). We come back to the approximation problem afterwards.

3. COMPUTATION OF THE LOUDSPEAKER WEIGHTS

To deploy a continuous loudspeaker is in general unpractical. We can instead approximate the aperture function of this continuous loudspeaker using an array of point-loudspeakers. To this purpose, we first low-pass and sample the continuous aperture function $s(\mathbf{x}_s, \omega)$ that can take values anywhere in the room. The gain of a linear loudspeaker array is then calculated by computing the least-squares approximation between the sampled aperture function and the (sampled) constrained function for the linear spatial manifold (10).

The sampling of $s(\mathbf{x}_s, \omega)$ induces a periodic extension on the set of Fourier coefficients $\beta_{\mathbf{q}}(\omega)$. Let Γ denote the sampling lattice that defines the spatial samples of the aperture function $s(\mathbf{x}_s, \omega)$ and assume $\Lambda \subseteq \Gamma$. Further let Σ be a sub-lattice of the spatial-frequency sampling lattice Φ i.e., $\Sigma \subseteq \Phi$, denoting the spatial-frequency periodicity lattice. The generator matrices are then related as $\mathbf{\Gamma} = 2\pi \mathbf{\Sigma}^{-\mathsf{T}}$. The samples of the aperture function are thus approximated by [7]

$$s(\mathbf{\Gamma n}, \omega) \approx s_{\mathrm{lp}}(\mathbf{\Gamma n}, \omega) = \frac{1}{\mathrm{N}(\Lambda/\Gamma)|\mathbf{\Gamma}|} \sum_{\mathbf{q} \in \mathcal{N}} \beta_{\mathbf{q}}(\omega) \mathrm{e}^{j\mathbf{q}^\mathsf{T} \mathbf{\Phi}^\mathsf{T} \mathbf{\Gamma n}}, \quad (11)$$

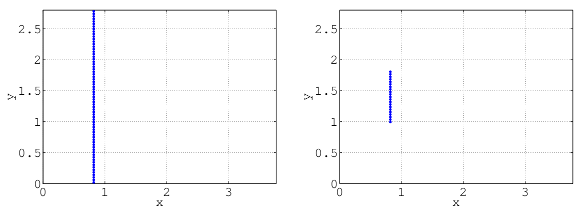
where $\mathbf{n} = [n_x, n_y]^\mathsf{T}$ is an integer vector, $\mathbf{\Gamma n} \in V_\Lambda(\mathbf{0})$, and $\mathrm{N}(\Lambda/\Gamma)$ denotes the number of lattice points of Γ that lie inside $V_\Lambda(\mathbf{0})$. The summation is now restricted to that set of lattice points: $\mathcal{N} = \{\mathbf{q} : \mathbf{\Phi q} \in V_\Sigma(\mathbf{0})\}$. It follows that $s_{\mathrm{lp}}(\mathbf{\Gamma n}, \omega)$ represents the samples of a spatially low-passed approximation of the aperture function $s(\mathbf{\Gamma n}, \omega)$. Clearly, the larger we make $\mathcal{N}$ the less low-passed this approximation will be. Moreover, note that the summation on the right hand side of (11) defines a 2D inverse DFT (IDFT).

As mentioned before, the spatial sampling lattice Γ is the reciprocal of the spatial-frequency periodicity lattice Σ, the generator matrices are related as $\mathbf{\Gamma} = 2\pi\mathbf{\Sigma}^{-\mathsf{T}}$. On the other hand, lattice Σ is a sub-lattice of the spatial-frequency sampling lattice Φ. The generator matrices are therefore related as $\mathbf{\Sigma} = \mathbf{\Phi N}$, where $\mathbf{N}$ is a full-rank *integer* matrix. Using these relationships the argument of the exponential function in (11) is given by, $\mathbf{q}^\mathsf{T}\mathbf{\Phi}^\mathsf{T}\mathbf{\Gamma n} = 2\pi\mathbf{q}^\mathsf{T}\mathbf{N}^{-\mathsf{T}}\mathbf{n}$. The IDFT (11) can therefore be rewritten as,

$$s_{\mathrm{lp}}(\mathbf{\Gamma n}, \omega) = \frac{1}{|\mathbf{N}||\mathbf{\Gamma}|} \sum_{\mathbf{q} \in \mathcal{N}} \beta_{\mathbf{q}}(\omega) \mathrm{e}^{j2\pi\mathbf{q}^\mathsf{T}\mathbf{N}^{-\mathsf{T}}\mathbf{n}}. \quad (12)$$

One can verify that $|\mathbf{N}| = \mathrm{N}(\Lambda/\Gamma) = \mathrm{N}(\Sigma/\Phi)$. In order to express this IDFT in matrix form, let us introduce the multi-dimensional DFT matrix of size $|\mathbf{N}| \times |\mathbf{N}|$, $\mathbf{F_N}$ with entries given by, $[\mathbf{F_N}]_{l,m} = \mathrm{e}^{-j2\pi\mathbf{n}_l^\mathsf{T}\mathbf{N}^{-1}\mathbf{q}_m}$, for $\mathbf{n} \in V_{\mathbf{N}^\mathsf{T}}(\mathbf{0})$ and $\mathbf{q} \in V_{\mathbf{N}}(\mathbf{0})$, where l and m denote an ordering of the vectors $\mathbf{n} \in V_{\mathbf{N}^\mathsf{T}}(\mathbf{0})$ and $\mathbf{q} \in V_{\mathbf{N}}(\mathbf{0})$ respectively. Matrix $\mathbf{F_N}$ is unitary satisfying $(\mathbf{F_N})^\mathsf{H}\mathbf{F_N} = |\mathbf{N}|\mathbf{I}$ and $(\mathbf{F_N})^{-1} = (\mathbf{F_N})^\mathsf{H}/|\mathbf{N}|$. For simplicity, let us make $\mathbf{N} = \mathrm{diag}(N_x, N_y)$. This means that $\mathbf{\Sigma} = \mathbf{\Phi N}$ is diagonal (since $\mathbf{\Phi}$ is diagonal in this work), and implies a rectangular packing of the spatial-frequency space. Using this notation, (12) is rewritten as

$$\mathbf{s}(\omega) = \frac{(\mathbf{F_N})^\mathsf{H}\boldsymbol{\beta}(\omega)}{|\mathbf{N}||\mathbf{\Gamma}|} = \frac{(\mathbf{F}_{N_x} \otimes \mathbf{F}_{N_y})^\mathsf{H}\boldsymbol{\beta}(\omega)}{(N_xN_y)|\mathbf{\Gamma}|}, \quad (13)$$



(a) Linear loudspeaker array (b) Truncated linear array

Fig. 2: Configurations considered in the experiments.

where $\otimes$ is the Kronecker product, $\boldsymbol{\beta}(\omega)$ and $\mathbf{s}(\omega)$ are $|\mathbf{N}|\times 1$ vectors with elements ordered as $\mathbf{q}_m$ and $\mathbf{n}_l$ respectively. The $_{\text{lp}}$ subscript indicating a spatially low-passed sound field has been dropped for readability.

Let us come back to the approximation problem in the example given at the end of the previous section. The vector of spatial Fourier coefficients defining the continuous linear loudspeaker is constructed as $\bar{\boldsymbol{\beta}}(\omega)=\bar{\boldsymbol{\beta}}^{(x)}(\omega)\otimes\bar{\boldsymbol{\beta}}^{(y)}(\omega)$, where $\bar{\boldsymbol{\beta}}^{(x)}(\omega)$ is a $N_x\times 1$ vector of fixed values taken from (10), and $\bar{\boldsymbol{\beta}}^{(y)}(\omega)$ is a $N_y\times 1$ vector of unknown values to be computed. From (13), we now pose the minimization problem that leads to the least-squares approximation of the linear loudspeaker array that renders the desired sound field. Make $\mathbf{B}=\bar{\boldsymbol{\beta}}^{(x)}(\omega)\otimes\mathbf{I}_{N_y}$, with $\mathbf{I}_{N_y}$ the identity matrix of size $N_y\times N_y$. Then $\bar{\boldsymbol{\beta}}(\omega)=\mathbf{B}\bar{\boldsymbol{\beta}}^{(y)}(\omega)$ and the problem is

$$\min_{\bar{\boldsymbol{\beta}}^{(y)}}\|\mathbf{B}\bar{\boldsymbol{\beta}}^{(y)}(\omega)-\boldsymbol{\beta}(\omega)\|^2. \quad (14)$$

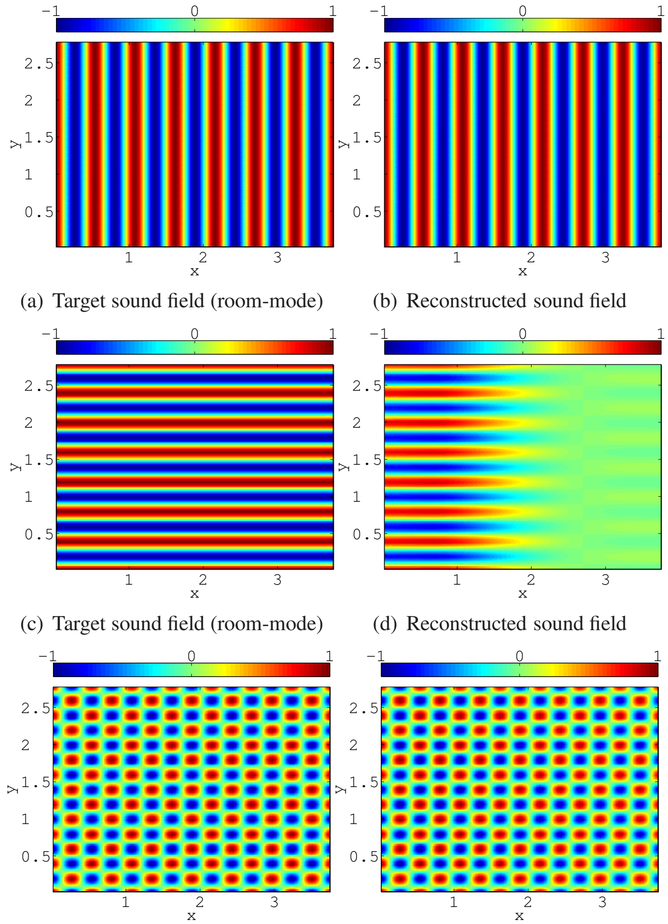
The optimal y-direction Fourier coefficients are then given by $\bar{\boldsymbol{\beta}}^{(y)}(\omega)=\mathbf{B}^{+}\boldsymbol{\beta}(\omega)$, where $\mathbf{B}^{+}$ is the pseudo-inverse of $\mathbf{B}$.

The loudspeaker weights that render the target sound field are computed by (12) or (13) using the calculated weights in (14). The computation of the weights is upper bounded by a complexity order $\mathcal{O}(N^3)$ (assuming $N=N_x\approx N_y$), but in many cases the structure in $\mathbf{B}$ might be exploited to perform the least-squares approximation more efficiently. The synthesis in (12) is performed using a IFFT with complexity order $\mathcal{O}(N^2\log N)$. This constitutes an efficient algorithm for sound field reproduction in box-shaped rooms.

4. SIMULATION RESULTS

We set as target sound-fields a set of steady-state room-modes at different temporal (and spatial) frequencies. We consider the ideal case of omnidirectional loudspeakers with flat frequency responses. We set one linear array covering the whole room space and for comparison, the case of a truncated linear array. The room dimensions are 3.81m in the x-direction and 2.84m in the y-direction.

We first address the case when the array fully covers the room length. The scenario is given in Fig. 2(a), The positions of the loudspeakers are depicted by little (blue) circles. The line runs parallel to the y-direction at $x=0.81$m. The linear array consists of 92 loudspeakers covering the full room length. The target sound field is given by a room-mode at temporal frequency $f=630$Hz and spatial frequencies $f_x=1.85$ cycles per meter and $f_y=0$ cycles per meter. The target sound field is thus given by a standing-wave with constant amplitude in the y-direction, which is depicted in Fig. 3(a). The sound-field synthesized by the calculated loudspeaker weights in (12) is depicted in 3(b). Since the room-mode displays surfaces of constant amplitude parallel to the loudspeaker array, it is perhaps not surprising that such a good reconstruction can be achieved.



(a) Target sound field (room-mode) (b) Reconstructed sound field

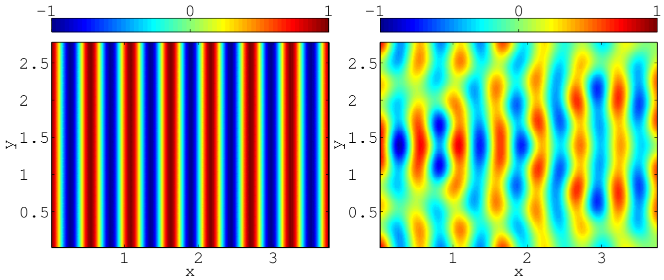
(c) Target sound field (room-mode) (d) Reconstructed sound field

(e) Target sound field (room-mode) (f) Reconstructed sound field

Fig. 3: Reconstruction of different room-modes using the full linear array.

To contrast, let us now try and reconstruct a target sound field with surfaces of constant amplitude perpendicular to the loudspeaker array. The target is a room-mode with frequencies $f=850$Hz, $f_x=0$ cycles/m and $f_y=2.5$ cycles/m. The target sound field is depicted in Fig. 3(c). The reconstructed sound field is given in Fig. 3(d).

It is clear that the linear array cannot reproduce this room-mode accurately in the whole room. This provides insight into the problem; to reconstruct a sound field composed of ill-reconstructed room-modes, one could compute another set of loudspeaker weights to better reconstruct the



(a) Target sound field (room-mode) (b) Reconstructed sound field

Fig. 4: Reconstruction of a room-mode using the truncated linear array.

sound field in a smaller zone, or properly weigh or discard those ill-reconstructed room-modes. Additionally, one could increase the degrees of freedom in the x-direction by placing more loudspeakers, for example, in a perpendicular line to form an "L"-shaped array.

As a third experiment, we set the target sound-field to a more general room-mode with frequencies $f = 850$Hz, $f_x = 1.85$ cycles/m and $f_y = 2.5$ cycles/m. The target is depicted in Fig. 3(c) and the reconstructed sound field is depicted in Fig. 3(f). The linear array can reconstruct this particular room-mode fairly accurately. Why this room-mode can be better reconstructed than the one in Fig. 3(d) has still to be analyzed. A guess would be by a more beneficial interference pattern of direct and reflected wavefronts. It is also physically hard for this linear loudspeaker array to sustain perpendicular planes of constant amplitude as distance increases.

For comparison, we also consider the reconstruction accuracy of the truncated linear array depicted in Fig. 2(b). In this case, only 21 loudspeakers are used covering a length from $y = 1$m to $y = 1.80$m. Truncation is characterized in space by a multiplication of the source function with a rectangular window. This translates into a (complex-valued) convolution in the spatial Fourier domain. From Fig. 3(d), it is seen that truncation has a clear impact in the performance. This is not surprising, since we are using the solution to the least-squares approximation for the full linear array. Naturally, it is also possible to solve the least squares problem including this windowing effect. The effect of the convolution in the spatial Fourier domain is, however, an important topic to be explored.

5. CONCLUSIONS

A new approach to sound field reproduction in box-shaped rooms is presented. We cover the efficient synthesis of room-modes using an array of loudspeakers, a topic not yet widely covered in the literature [3]. We have proposed an analytic method to model and match the room-modes in the spatio-temporal Fourier domain. We provided an example setup using a linear array parallel to one of the room walls, but the algorithm can be readily applied to other array geometries. The method gives insight into the problem since vanishing modes (components that cannot be inverted or equalized), can be identified and a strategy to handle them can be drawn accordingly. Our simulation results show that reproducible room-modes can be accurately and efficiently synthesized. The theory and application given in this work is restricted to box-shaped rooms with fully reflective walls. An extension to include wall absorption and arbitrary room geometries is a topic of current research.

REFERENCES

[1] A. J. Berkhout, D. de Vries, J. Baan, and B. W. van den Oetelaar, "A wave field extrapolation approach to acoustical modeling in enclosed spaces.," *J. Acoust. Soc. Am.*, vol. 105, no. 3, pp. 1725–1733, Mar. 1999.

[2] J. Ahrens and S. Spors, "Analytical driving functions for higher order ambisonics," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. - Apr. 2008, pp. 373 –376.

[3] J. Ahrens, *Analytic Methods of Sound Field Synthesis*, T-Labs Series in Telecommunication Services. Springer, Berlin, Jan. 2012.

[4] P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Inverse filter design and equalization zones in multichannel sound reproduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 3, no. 1, pp. 185 –192, Jan. 1995.

[5] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *J. Acoust. Soc. Am.*, vol. 117, no. 4, pp. 2100–2111, Apr. 2005.

[6] W. Jin and W. B. Kleijn, "Multizone soundfield reproduction in reverberant rooms using compressed sensing techniques," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2014, vol. 5, pp. 4728 – 4732.

[7] J. Martinez and R. Heusdens, "On low-complexity simulation of multichannel room impulse responses," *IEEE Signal Process. Lett.*, vol. 17, no. 7, pp. 667 –670, July 2010.

[8] Y. J. Wu and T. D. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 1, pp. 107 –116, jan. 2009.

[9] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[10] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling.," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.