A SPATIO-TEMPORAL GENERALIZED FOURIER DOMAIN FRAMEWORK TO ACOUSTIC MODELING IN ENCLOSED SPACES

Jorge Martinez, Richard Heusdens and Richard C. Hendriks

Signal & Information Processing Lab, Delft University of Technology, 2628 CD Delft, the Netherlands
{J.A.MartinezCastaneda,R.Heusdens,R.C.Hendriks}@TuDelft.nl

ABSTRACT

In this paper, we present a spatio-temporal framework for multichannel acoustic modeling in enclosed spaces. Reverberation occurs when the sound field is enclosed between reflective boundaries (e.g. walls). We model the reverberated sound field by proper sampling of the (generalized) Fourier representation of the free-field sound field. We show that the spatial aliasing introduced by spectral sampling represents all the (damped) reflections. From the samples of the generalized spectrum, we compute the spatio-temporal sound field in the enclosed space with very low-complexity, of $\mathcal{O}(N \log N)$ per measuring position, with N proportional to the reverberation time.

***Index Terms*—** Room impulse response, generalized Fourier transform, spatio-temporal processing.

1. INTRODUCTION

In many situations, human acoustic communication occurs in enclosed spaces. In multichannel telecommunication technologies, e.g., hands-free devices and teleconference systems, the acoustical properties of the enclosed space become more relevant than in the single-channel case, since reverberation and acoustic echo feedback become greater challenges [1]. Moreover, near future technologies that range from immersive telegaming to telepresence conference systems, and medical systems such as telesurgery, will require even greater degrees of acoustic control over hundreds of even thousands of acoustic input/output channels [2]. It is clear that to enable real-time, full-duplex telecommunications for such systems, new fast, scalable signal processing approaches for estimation and modeling of the spatio-temporal sound field are needed.

In this paper we present a fast, spatio-temporal framework for multichannel acoustical modeling in enclosed spaces. In Fourier analysis, it is well known that sampling in one domain implies periodicity in the reciprocal domain [3]. In [4], this property is exploited, and a fundamental relationship between the free-field sound field, and the sound field in a room with fully reflective walls is made. In this case, the reflections can be modeled by a perfectly periodic structure, or equivalently, by a carefully chosen sampling condition on the 4-D spectrum of the sound field [4]. In this paper we further extend the approach to include constant wall reflection coefficients. First, we show how the enclosed sound field, as modeled by the mirror image source method (MISM) [5], can be expressed as a geometrically weighted periodic summation. Then we associate this infinite summation, with a sampling condition in a generalized Fourier domain. The Poisson summation formula for the generalized Fourier transform (GFT) [6], relates the geometrically weighted summation of the sound field over a lattice, to the samples of its generalized spectrum over the reciprocal lattice. From these samples, we model with very low complexity ($\mathcal{O}(N \log N)$ in the reverberation time per channel) the sound field at every measuring position.

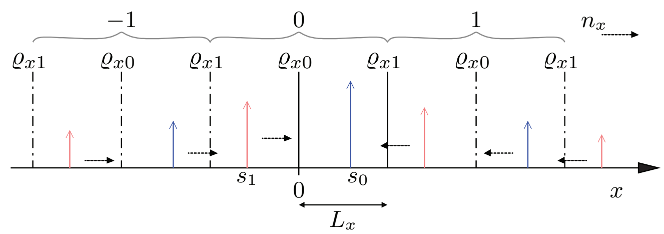

Fig. 1: *Example of the damped repetitions of the free-field sound field used in the 1-D MISM, to construct the reverberated sound field.*

2. THE SOUND FIELD IN AN ENCLOSED SPACE

In order to simplify the discussion, let us first analyze the sound field as modeled by the MISM in 1-D, and later extend the results to higher dimensions. We set the first boundary (wall) at position $x = 0$, and denote by $0 \leq \varrho_{x0} \leq 1$ the associated reflection coefficient. The second boundary is set at a distance L_x, its reflection coefficient is denoted by $0 \leq \varrho_{x1} \leq 1$. The scenario is depicted in Fig. 1, where the boundaries are indicated by solid lines. Given a free-field sound field, $p_0(x,t)$, the MISM models the reverberated sound field inside the enclosure (the interval $0 \leq x \leq L_x$ in this case), as the contribution of infinitely many, geometrically damped spatial repetitions of $p_0(x,t)$. In general $p_0(x,t)$ consist of two, full bandwidth functions $p_{0+}(x-ct)$ and $p_{0-}(x+ct)$ (the d'Alembert solution [7]), having compact support (shorter than L_x) in the interval $0 \leq x \leq L_x$, traveling to the right and left respectively.

Let $p_{0+}(x) = p_{0-}(x) = \delta(x - s_0)$, this is, a Dirac's delta pulse located at position $x = s_0$. To correctly account for all reflections, a second function is used in the MISM, $p_1(x,t) = p_0(-x,t)$, a spatial-reversed version of the free-field sound field, starting at position $x = s_1 = -s_0$. The reverberated sound field $p(x,t)$, is obtained by a superposition of infinitely many geometrically damped periodic repetitions of $p_0(x,t)$ and $p_1(x,t)$. We thus call these two functions the "mother" sources. Let $\varrho_x = \varrho_{x0}\varrho_{x1}$. We then obtain,

$$\begin{aligned} p(x,t) = &\sum_{n_x \in \mathbb{Z}} \varrho_x^{|n_x|}\, p_{0+}(x + (2L_x)n_x, t) \\ &+ \sum_{n_x \in \mathbb{Z}} \varrho_x^{|n_x|}\, p_{0-}(x + (2L_x)n_x, t) \\ &+ \varrho_{x0} \sum_{n_x \in \mathbb{Z}} \varrho_x^{|n_x|}\, p_{1+}(x + (2L_x)n_x, t) \\ &+ \varrho_{x0}^{-1} \sum_{n_x \in \mathbb{Z}} \varrho_x^{|n_x|}\, p_{1-}(x + (2L_x)n_x, t), \end{aligned}$$

where $p_{0+}(x,t) = p_0(x,t)H(x - s_0)$, $p_{0-}(x,t) = p_0(x,t)H(-x + s_0)$, $p_{1+}(x,t) = p_1(x,t)H(x - s_1)$, and $p_{1-}(x,t) = p_1(x,t)H(-x + s_1)$, with $H(x)$ the unit-step function. Note that for p_1 and its repetitions, the "alignment" factors $\varrho_{x0}^{\pm 1}$ are necessary to obtain the correct damping effect inside the enclosure.

Next, we extend this model to 3-D in space, including the inverse-square law of sound propagation. In this case, the free-field sound field p_0, as it is registered at position $\mathbf{x}=[x,y,z]^T$, and time t is given by the plenacoustic function (PAF) [8], i.e. the free-field Green's function [7],

$$p_0(\mathbf{x},t) = \frac{\delta\left(t-\frac{r}{c}\right)}{4\pi r}, \qquad (1)$$

where $r=\|\mathbf{x}-\mathbf{s}_0\|$ is the distance from measurement point $\mathbf{x}$ to the position of the source at time $t=0$, i.e. $\mathbf{s}_0=[x_{s0},y_{s0},z_{s0}]^T$. As in the 1-D case, the MISM models the sound field inside the room by an infinite number of repetitions of the mother sources in every coordinate. Therefore in general, 2^ν mother sources are needed for the ν-D scenario [4]. The full set of repetitions is then obtained by a 3-D periodic packing over a lattice, say Λ. We analyze the case of box-shaped rooms, so that the generator matrix of Λ is given by $\mathbf{\Lambda} = \mathrm{diag}\,(2L_x, 2L_y, 2L_z)$ [4]. As in the 1-D case, we need to consider waves traveling exclusively to the positive and negative x, y and z directions. To do this we let O_q, $q=0,\dots,7$, denote the qth octant of the 3-D space, and use a binary ordering scheme based on the signs of the coordinates to enumerate the octants. This is, $(x\geq 0, y\geq 0, z\geq 0)=(+,+,+)=0$, $(x<0, y\geq 0, z\geq 0)=(-,+,+)=1$, and so on. Then for example, $\mathrm{O}_3 \triangleq \{\mathbf{x}\in\mathbb{R}^3 : x<0, y<0, z\geq 0\}$. Let $H_q(\mathbf{x})$ denote the 3-D unit-step function given by

$$H_q(\mathbf{x}) \triangleq \begin{cases} 1, & \mathbf{x}\in \mathrm{O}_q, \\ 0, & \text{otherwise.}\end{cases}$$

Then we define $p_{lq}(\mathbf{x},t) = p_l(\mathbf{x},t)H_q(\mathbf{x}-\mathbf{s}_l)$. The functions $p_{lq}(\mathbf{x},t)$ represent each of the eight causal and anticausal parts of the l-th mother source sound field $p_l(\mathbf{x},t)$. Let ϱ_{i0} for $i\in\{x,y,z\}$, denote the reflection coefficients of the boundaries adjacent to the origin in each direction, and ϱ_{i1} denote the reflection coefficients of the opposite boundaries. Let $\varrho_x=\varrho_{x0}\varrho_{x1}$, and define in the same way ϱ_y and ϱ_z. The reverberated sound field in the room is thus obtained by

$$p(\mathbf{x},t) = \sum_{l=0}^{7}\sum_{q=0}^{7}\varrho_{lq}\sum_{\mathbf{n}\in\mathbb{Z}^3}\left(\prod_i \varrho_i^{|n_i|}\right)p_{lq}(\mathbf{x}+\mathbf{\Lambda n},t), \qquad (2)$$

where $i\in\{x,y,z\}$ and $\mathbf{n}=[n_x,n_y,n_z]^T$. The corresponding alignment factors ϱ_{lq}, for each of the functions $p_{lq}(\mathbf{x},t)$, can be derived by extending the analysis given above for the 1-D case.

3. A GENERALIZED POISSON SUMMATION FORMULA

The classical Poisson summation formula relates a signal say, $p(\mathbf{x})\in L^2(\mathbb{R}^\nu)$, to the samples of its spectrum $P(\boldsymbol{\phi})$, as follows,

$$\sum_{\mathbf{n}\in\mathbb{Z}^\nu} p(\mathbf{x}+\mathbf{\Lambda n}) = \frac{1}{|\mathbf{\Lambda}|}\sum_{\mathbf{k}\in\mathbb{Z}^\nu} P(\mathbf{\Phi k})e^{j(\mathbf{k}^T\mathbf{\Phi}^T\mathbf{x})}\mathrm{d}\mathbf{x}, \qquad (3)$$

where $\mathbf{\Lambda}$ is a generating matrix for the lattice Λ, and $\mathbf{\Phi}=2\pi\mathbf{\Lambda}^{-T}$ is the generator matrix of the spectral sampling lattice Φ.

A generalized Poisson summation formula is introduced in [6], where the theory is presented for one dimensional functions. This is extended to multidimensional spaces as follows. For a signal $p(\mathbf{x})\in L^2(\mathbb{R}^\nu)$, we define the generalized Fourier transform (GFT) with parameter $\boldsymbol{\alpha}\in\mathbb{C}^\nu : \alpha_i\neq 0, \forall i=0,\dots,\nu-1$, as

$$\mathcal{F}_{\boldsymbol{\alpha}}\left\{p(\mathbf{x})\right\} \triangleq P_{\boldsymbol{\alpha}}(\boldsymbol{\phi}) = \int_{\mathbb{R}^\nu} p(\mathbf{x})e^{\boldsymbol{\beta}^T\mathbf{x}}e^{-j(\boldsymbol{\phi}^T\mathbf{x})}\mathrm{d}\mathbf{x}, \qquad (4)$$

where $\boldsymbol{\beta}=\mathbf{\Lambda}^{-T}\log(\boldsymbol{\alpha})$, $\mathbf{\Lambda}^{-T}$ is the generator matrix of the reciprocal lattice of Λ, and $\log(\boldsymbol{\alpha})\triangleq[\log(\alpha_0),\dots,\log(\alpha_{\nu-1})]^T$. The inverse transform is given by,

$$\mathcal{F}_{\boldsymbol{\alpha}}^{-1}\left\{P_{\boldsymbol{\alpha}}(\boldsymbol{\phi})\right\} \triangleq p(\mathbf{x}) = \frac{e^{-\boldsymbol{\beta}^T\mathbf{x}}}{(2\pi)^\nu}\int_{\mathbb{R}^\nu} P_{\boldsymbol{\alpha}}(\boldsymbol{\phi})e^{j(\mathbf{x}^T\boldsymbol{\phi})}\mathrm{d}\mathbf{x}. \qquad (5)$$

The GFT denoted by (4) is equivalent to the Fourier transform (FT) (if it can be defined) of the modulated signal $p(\mathbf{x})e^{\boldsymbol{\beta}^T\mathbf{x}}$ [6]. When $\boldsymbol{\alpha}=[1,1,\dots,1]^T$, the transform pair (4) and (5) corresponds to the standard FT pair. The generalized Poisson summation formula follows by evaluating (4) in (3), i.e.

$$\sum_{\mathbf{n}\in\mathbb{Z}^\nu} e^{\boldsymbol{\beta}^T\mathbf{\Lambda n}}p(\mathbf{x}+\mathbf{\Lambda n}) = \frac{e^{-\boldsymbol{\beta}^T\mathbf{x}}}{|\mathbf{\Lambda}|}\sum_{\mathbf{k}\in\mathbb{Z}^\nu} P_{\boldsymbol{\alpha}}(\mathbf{\Phi k})e^{j(\mathbf{k}^T\mathbf{\Phi}^T\mathbf{x})}. \qquad (6)$$

Since $\boldsymbol{\beta}^T\mathbf{\Lambda n}=\log(\boldsymbol{\alpha})^T\mathbf{\Lambda}^{-1}\mathbf{\Lambda n}=\sum_{i=0}^{\nu-1} n_i\log\alpha_i$, then we make

$$e^{\log(\alpha)^T\mathbf{n}} = \exp\left(\sum_{i=0}^{\nu-1} n_i\log\alpha_i\right) = \prod_{i=0}^{\nu-1}\alpha_i^{n_i},$$

so that (6) can be rewritten as,

$$\sum_{\mathbf{n}\in\mathbb{Z}^\nu}\left(\prod_{i=0}^{\nu-1}\alpha_i^{n_i}\right)p(\mathbf{x}+\mathbf{\Lambda n}) = \frac{e^{-\boldsymbol{\beta}^T\mathbf{x}}}{|\mathbf{\Lambda}|}\sum_{\mathbf{k}\in\mathbb{Z}^\nu} P_{\boldsymbol{\alpha}}(\mathbf{\Phi k})e^{j(\mathbf{k}^T\mathbf{\Phi}^T\mathbf{x})}. \qquad (7)$$

Let us define $\tilde{p}_{\boldsymbol{\alpha}}(\mathbf{x})\triangleq\sum_{\mathbf{n}\in\mathbb{Z}^\nu}\left(\prod_i\alpha_i^{n_i}\right)p(\mathbf{x}+\mathbf{\Lambda n})$. It follows that (7) relates a geometrically weighted periodic extension of the signal $\tilde{p}_{\boldsymbol{\alpha}}(\mathbf{x})$ over a lattice Λ, to the samples of its generalized spectrum $P_{\boldsymbol{\alpha}}(\mathbf{\Phi k})$ over the (scaled) reciprocal lattice Φ.

If $\boldsymbol{\beta}\in\mathbb{C}^\nu : |\Re\{\beta_i\}|=0,\ \forall\beta_i$, is a vector of purely imaginary components, then $j\boldsymbol{\beta}$ is a vector of purely real components and, from the modulation theorem for the FT [3] we have that,

$$P_{\boldsymbol{\alpha}}(\boldsymbol{\phi}) \triangleq \mathcal{F}\left\{p(\mathbf{x})e^{\boldsymbol{\beta}^T\mathbf{x}}\right\} = P(\boldsymbol{\phi}+j\boldsymbol{\beta}). \qquad (8)$$

In the next section, we exploit this fact to obtain the generalized Fourier spectrum from the standard spectrum of the sound field.

4. FAST MODELING OF THE ENCLOSED SOUND FIELD

In this section we present a fast method to compute the full sound field in a room. We cannot directly associate the sound field as obtained by the geometrically damped summation in (2), to a sampling condition in the generalized Fourier domain. In order to do that, we first show that (2) can be expressed as a geometrically weighted periodic summation (in the form given by (7)), and then we associate this summation to the samples of the generalized spectra of the sound fields $p_{lq}(\mathbf{x},t)$. We first define $\varsigma_q(x)=\mathrm{sign}(x)$, $\varsigma_q(y)=\mathrm{sign}(y)$ and $\varsigma_q(z)=\mathrm{sign}(z)$ for $(x,y,z)\in\mathrm{O}_q$. Clearly, for $i\in\{x,y,z\}$ each $\varsigma_q(i)=\pm 1$ depending on the signs of the coordinates defining the octant O_q, then we write,

$$p(\mathbf{x},t) = \sum_{l=0}^{7}\sum_{q=0}^{7}\varrho_{lq}\sum_{\mathbf{n}\in\mathbb{Z}^3}\left(\prod_i\varrho_i^{\varsigma_q(i)n_i}\right)p_{lq}(\mathbf{x}+\mathbf{\Lambda n},t),\quad \mathbf{x}\in V_\Lambda(\mathbf{0}), \qquad (9)$$

where $V_\Lambda(\mathbf{0})$ is the (central) Voronoi region of the lattice Λ at the origin. The geometrically weighted periodic summation in (9), gives exactly the same sound field as (2) for $\mathbf{x}\in V_\Lambda(\mathbf{0})$ (which includes the actual room space). To see this, we set for a moment $n_y=n_z=0$. Then in the zone $|x-x_{s_l}|\leq 2L_x$ or $|x|\leq L_x$, we have that,

$$\varrho_x^{|n_x|}p_l(\mathbf{x}+\mathbf{\Lambda n},t)H_q(\mathbf{x}-\mathbf{s}_l+\mathbf{\Lambda n})=0,$$

for $n_x<0$ and q even (i.e. $\varsigma_q(x)=1$), and in the same way for $n_x>0$, and q uneven (i.e. $\varsigma_q(x)=-1$), per definition of the Heaviside functions $H_q(\mathbf{x})$. Therefore,

$$\varrho_x^{|n_x|}p_{lq}(\mathbf{x}+\mathbf{\Lambda n},t)=\varrho_x^{\varsigma_q(x)n_x}p_{lq}(\mathbf{x}+\mathbf{\Lambda n},t),\quad |x|\leq L_x.$$

An equivalent situation occurs in the y and z directions, for $n_y \neq 0$ and $n_z \neq 0$ respectively, so (9) equals (2) in the zone $\mathbf{x} \in V_\Lambda(\mathbf{0})$. Next, we have the following result,

Proposition 1 *Let $\mathbf{\Lambda}$ be the generator matrix of the lattice specifying the spatial periodic packing of the sound fields $p_{lq}(\mathbf{x}, t)$, and let $\mathbf{\Phi}$ be the matrix basis of the lattice specifying the sampling points of the spatial-generalized spectra. If $\mathbf{\Phi} = 2\pi\mathbf{\Lambda}^{-T}$, then the functions $P_{\boldsymbol{\alpha}lq}(\mathbf{\Phi k}, \omega)$, $\mathbf{k} \in \mathbb{Z}^3$, are the generalized Fourier coefficients of*

$$\sum_{\mathbf{n}\in\mathbb{Z}^3} \left(\prod_i \varrho_i^{\varsigma_q(i) n_i}\right) p_{lq}(\mathbf{x} + \mathbf{\Lambda n}, t), \tag{10}$$

if and only if $\boldsymbol{\alpha} = [\varrho_x^{\varsigma_q(x)}, \varrho_y^{\varsigma_q(y)}, \varrho_z^{\varsigma_q(z)}, 1]^T$.

The proof follows immediately from the generalized Poisson summation formula, with $\boldsymbol{\alpha} = [\varrho_x^{\varsigma_q(x)}, \varrho_y^{\varsigma_q(y)}, \varrho_z^{\varsigma_q(z)}, 1]^T$, the parameter of the 4-D spatio-temporal GFT.

This result gives us a formula for reconstructing the full sound field in a room given by (9). Sample the generalized spectra $P_{\boldsymbol{\alpha}lq}(\boldsymbol{\phi}, \omega)$ using the sampling lattice generated by $\mathbf{\Phi} = 2\pi\mathbf{\Lambda}^{-T}$, apply a temporal inverse GFT with parameter $\alpha = 1$ (i.e. an inverse FT), then use the generalized Poisson summation formula given by (7), on the coefficients just obtained to synthesize the functions (10), which are then used to calculate the reverberated sound field as in (9). For a box-shaped room with dimensions (L_x, L_y, L_z), the generator matrix for the spectral sampling lattice is thus defined by, $\mathbf{\Phi} = 2\pi\mathbf{\Lambda}^{-T} = \mathrm{diag}\,(\pi/L_x, \pi/L_y, \pi/L_z)$ [4].

To implement the proposed algorithm on a digital computer, we need to sample the temporal frequency ω as well. This will introduce undesired temporal aliasing, since $p(\mathbf{x}, t)$ is clearly not time limited (it has infinite support). Let $\mathbf{\Psi}$ denote the generator matrix for the lattice specifying the sampling points of both the spatial and temporal frequency variables, defined by $\mathbf{\Psi} = \mathrm{diag}\,(\mathbf{\Phi}, \Omega_s)$, where Ω_s denotes the temporal-frequency sampling interval. The diagonal form of $\mathbf{\Psi}$ implies independent sampling of spatial and temporal frequencies. Then, from the generalized Poisson summation formula and the arguments in Proposition 1, it follows that the sampled spectra $P_{\boldsymbol{\alpha}lq}(\mathbf{\Psi k})$, $\mathbf{k} \in \mathbb{Z}^4$, are the generalized Fourier coefficients of

$$\tilde{p}_{\boldsymbol{\alpha}lq}(\mathbf{x}, t) = \sum_{\mathbf{n}\in\mathbb{Z}^3}\sum_{n\in\mathbb{Z}} \left(\prod_i \varrho_i^{\varsigma_q(i) n_i}\right) p_{lq}(\mathbf{x} + \mathbf{\Lambda n}, t + T_s n).$$

if $\mathbf{\Psi} = 2\pi\mathbf{\Delta}^{-T} = 2\pi\,\mathrm{diag}\,(\mathbf{\Lambda}, T_s)^{-T}$. This last equality defines $\mathbf{\Delta}$, where $T_s = 2\pi/\Omega_s$ is the interval of temporal periodicity. Clearly, the summation over $n \in \mathbb{Z}$, $n \neq 0$, represents the time-domain aliasing introduced by sampling the temporal frequency ω.

Although the sound field $p_0(\mathbf{x}, t)$ given by (1) has infinite temporal support, we have that $\lim_{t\to\infty} p_0(\mathbf{x}, t) = 0$ for $\mathbf{x} \in V_\Lambda(\mathbf{0})$, so that by making T_s sufficiently large (i.e. making Ω_s sufficiently small), we can make the error due to time-domain aliasing neglectable. Taking this into account, the total sound field can be approximated by

$$p(\mathbf{x}, t) \approx \sum_{l=0}^{7}\sum_{q=0}^{7} \frac{e^{-\boldsymbol{\beta}_q^T \mathbf{x} - \beta t}}{|\mathbf{\Delta}|} \sum_{\mathbf{k}\in\mathbb{Z}^4} \varrho_{lq} P_{\boldsymbol{\alpha}lq}(\mathbf{\Psi k}) e^{j(\mathbf{k}^T \mathbf{\Psi}^T [\mathbf{x}^T, t]^T)}, \tag{11}$$

where $\boldsymbol{\beta}_q = \mathbf{\Lambda}^{-T}[\log(\varrho_x^{\varsigma_q(x)}), \log(\varrho_y^{\varsigma_q(y)}), \log(\varrho_z^{\varsigma_q(z)})]^T$, $\beta = \log(1)/T_s = 0$. The reconstruction of $p(\mathbf{x}, t)$ out of its generalized Fourier coefficients involves an infinite summation. In addition, to get the sound field at another location in space we have to recompute (11). This computational complexity is clearly too high. To reduce the complexity, let us now limit the summation over $\mathbf{k} \in \mathbb{Z}^4$ in (11) to a finite number of elements and periodically extend this finite set over the 4-D spatio-temporal frequency space. By making the (sampled) generalized spectra periodic, we are imposing a discretization on the sound field. Let Σ be a sublattice of Ψ, denoting the lattice for generating the periodic packing of the frequency space. Next, let Γ denote the spatio-temporal sampling lattice imposed by making the generalized spectra periodic and assume $\Delta \subseteq \Gamma$. Then clearly we have that $\mathbf{\Gamma} = 2\pi\mathbf{\Sigma}^{-T}$. The functions $\tilde{p}_{\boldsymbol{\alpha}lq}$ can be calculated by,

$$\tilde{p}_{\boldsymbol{\alpha}lq}(\mathbf{\Gamma n}) = \frac{e^{-\boldsymbol{\beta}^T \mathbf{\Gamma n}}}{N(\Delta/\Gamma)} \sum_{\mathbf{k}\in V_{\mathbf{\Sigma}}(\mathbf{0})} |\mathbf{\Delta}|^{-1} P_{\boldsymbol{\alpha}lq}(\mathbf{\Psi k}) e^{j(\mathbf{k}^T \mathbf{\Psi}^T \mathbf{\Gamma n})} \tag{12}$$

where $\boldsymbol{\beta} = \mathbf{\Delta}^{-T}\log(\boldsymbol{\alpha})$. From here we see that by making $V_{\mathbf{\Sigma}}(\mathbf{0})$ larger (i.e. taking more frequency samples), the finer we sample the functions $\tilde{p}_{\boldsymbol{\alpha}lq}$. Certainly we have that,

$$N(\Delta/\Gamma) = \frac{|\mathbf{\Delta}|}{|\mathbf{\Gamma}|} = \frac{(2\pi)^4|\mathbf{\Psi}^{-T}|}{(2\pi)^4|\mathbf{\Sigma}^{-T}|} = \frac{|\mathbf{\Sigma}|}{|\mathbf{\Psi}|} = N(\Sigma/\Psi)$$

so that the number of evaluation points $N(\Sigma/\Psi)$ of the generalized spectra, equals the number of spatio-temporal samples of the functions $\tilde{p}_{\boldsymbol{\alpha}lq}$. Since the functions $\tilde{p}_{\boldsymbol{\alpha}lq}$ are of a geometrically weighted periodic form, we only need to evaluate them in $V_\Delta(\mathbf{0})$. The sampled sound field is thus obtained by,

$$p(\mathbf{\Gamma n}) \approx \sum_{l=0}^{7}\sum_{q=0}^{7} \varrho_{lq}\tilde{p}_{\boldsymbol{\alpha}lq}(\mathbf{\Gamma n}), \quad \text{for } \mathbf{\Gamma n} \in V_\Delta(\mathbf{0}). \tag{13}$$

The computational complexity is given by the evaluation of (12), limited to $\{\mathbf{n} \in \mathbb{Z}^4 : \mathbf{\Gamma n} \in V_\Delta(\mathbf{0})\}$. In [8], it is shown that the spectrum has much of its energy concentrated in the region $\|\boldsymbol{\phi}\| \leq |\omega/c|$. Then, for a maximum temporal frequency ω_b, this fact is used to determine $\mathbf{\Sigma}$, and therefore $N(\mathbf{\Sigma}/\mathbf{\Psi})$, allowing us for a trade between speed and accuracy of reconstruction. A careful examination of (12) reveals that the summation over $\mathbf{k}$ represents a standard discrete Fourier transform (DFT). The complexity is drastically reduced if the fast Fourier transform (FFT) is used instead. In this case, the operation will take only $\mathcal{O}(N^4 \log N)$ operations for computing (12) in $N(\Delta/\Gamma)$ spatio-temporal positions. Recognizing that $N(\Delta/\Gamma) = N(\Sigma/\Psi)$, then we can conclude that the method is of complexity $\mathcal{O}(N \log N)$ per receiver position.

5. EXPERIMENTAL RESULTS

In this section we compare the proposed approach against MISM [5]. The experiments are performed in MATLAB®, on a PC computer using a single CPU running at 2.4 GHz. The newly proposed method is implemented as a native m-file function and the MISM as a C++ mex-function. The computed scenario is displayed in Fig. 2(a). The room dimensions are $[L_x, L_y, L_z]^T = [4.78, 3.79, 2.83]^T$. The wall reflection coefficients are $\varrho_{i0} = 1$, and $\varrho_{i1} = -1$ for $i \in \{y, z\}$, and $\varrho_{x0} = \varrho_{x1} = 0$, so that $\varrho_y = \varrho_z = -1$, and $\varrho_x = 0$. We consider temporal signals bandlimited to 4KHz. The temporal sampling frequency is thus set to $\omega_s = (8000)2\pi$. To avoid the non-banlimited representation of the delta pulses in the MISM, we replace each delta pulse with the impulse response of a (8ms long) Hanning-windowed ideal low-pass filter with cut-off frequency set to the Nyquist frequency [9]. The post-processing high-pass filter, suggested in [5] is disabled in the experiments to avoid biased results. The simulation length is $T_h = 2.058$s, or equivalently in samples, $N_t = 16384$. We choose the temporal-frequency sampling interval to be $\Omega_s = \omega_s/2N_t$. The GFT parameter is set to $\boldsymbol{\alpha} = [\varrho_y^{\varsigma_q(y)}, \varrho_z^{\varsigma_q(z)}, 1]^T = [-1, -1, 1]^T$ for all q, which is 3-D since no reflections are considered in the x direction. The spectral-sampling

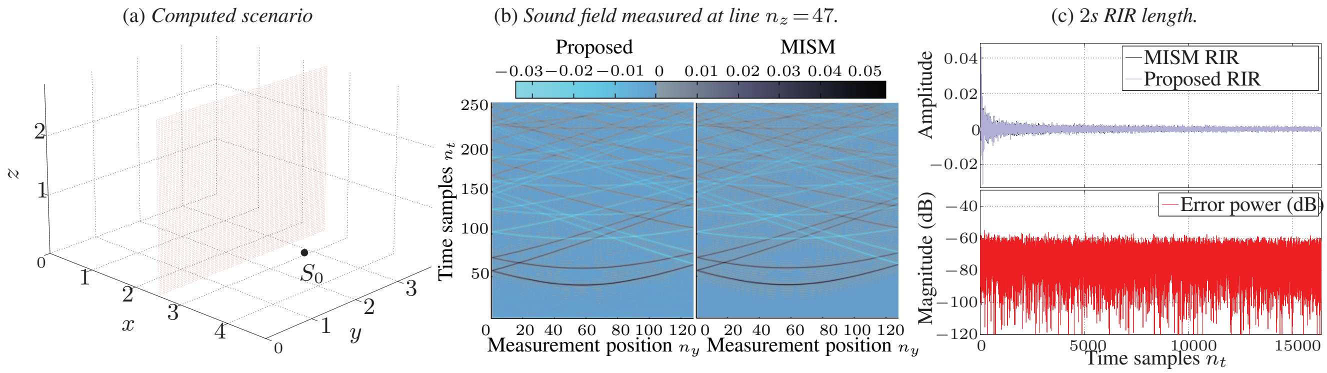

Fig. 2: *(a)* 12288 *measurement positions (*$128y$,$96z$*, red dots) at* $x = 2.51$. *The sound source is* S_0. *The wall reflection coefficients are* $\varrho_{i0} = 1$ *and* $\varrho_{i1} = -1$ *for* $i \in \{y, z\}$, *and* $\varrho_{x0} = \varrho_{x1} = 0$. *The proposed method took* 7 *minutes to complete, the MISM* 3 *days. (b) Experimental results for only one line in the* y *direction at* $n_z = 47$. *(c) Comparison of both methods only for receiver* $n_y = 55$ *of the same line.*

matrix is thus given by $\mathbf{\Psi} = \text{diag}\,(\pi/L_y, \pi/L_z, \Omega_s)$. The spectral-periodicity matrix is chosen to be $\mathbf{\Sigma} = 2\mathbf{\Psi}\,\text{diag}\,(N_y, N_z, N_t)$, with $N_y = 128$ and $N_z = 96$, which gives a good compromise between temporal aliasing and speed of reconstruction. This directly defines the spatio-temporal sampling lattice Γ for the sound field, with generator matrix $\mathbf{\Gamma}$=$\text{diag}\left(0.0296, 0.0294, 125 \times 10^{-6}\right)$. In this case, the diagonal form of $\mathbf{\Gamma}$ imposes a rectangular arrangement of the sound field measurement positions (spatial samples). This can be seen in Fig. 2(a), where the $N_y = 128$ and $N_z = 96$ measurement positions are arranged in a plane perpendicular to the x direction at $x = 2.51$m, giving a total of 12288 spatial positions to be calculated. The source S_0 position is $\mathbf{s}_0 = [3.98, 1.70, 0.63]^T$. Using (8), the generalized Fourier spectra are obtained from the 3-D spectrum of the PAF of the mother sources [8]. We evaluate the generalized spectrum of the mother sources at sampling positions defined by $\mathbf{\Psi k}$, to obtain the generalized Fourier coefficients. To synthesize the sound field given by (13) and (12), first a multidimensional inverse FFT is applied on the generalized Fourier coefficients, followed by a modulation by $e^{-\boldsymbol{\beta}^T\mathbf{\Gamma n}}$. Using the MISM algorithm, we modeled individually the RIRs from the source to each measuring position. We use a triplet of integers n_y, n_z, n_t, to index the sound field samples. The colormap plots in Fig. 2(b) show the results only for one measurement line in the y direction at $n_z = 47$, and time samples $0 \leq n_t \leq 256$. Additionally in Fig. 2(c), a plot is given where we compare the full, 2.058s in length RIRs, only for measurement position $n_y = 55$ for the same line. As it is seen, the spatio-temporal "locations" and amplitudes of the reflections are perfectly modeled by the newly proposed method. The discrepancies that are observed between both approaches, are caused by temporal aliasing, which can be made arbitrarily small by decreasing the spectral sampling interval Ω_s. The newly proposed method took 393s (or 6.5 minutes) to compute the full scenario, the MISM took 288072s (or 3.3 days) to complete, showing the contrasting difference in complexity.

6. CONCLUDING REMARKS

We derived a generalized Fourier series representation for enclosed sound fields that can be mathematically modeled in terms of a lattice structure. The representation is directly obtained by a sampling condition on the continuous generalized spectra of the free-field sound field. This leads to a fast method for multichannel modeling of the sound field inside the enclosure. We give results for box-shaped rooms where the generalized spectra are readily available, showing the huge advantage in computational complexity when compared against the MISM. To efficiently determine the generalized Fourier coefficients is however the topic of current research. The method can be extended to account for other room geometries, as long as the room space is defined by a convex polyhedron and the sound field can be modeled by a periodic packing. However, further research is needed to derive a general way to determine valid periodic packings of virtual sources given the desired room geometry (see e.g. [10]). The newly proposed framework is of special importance in the domain of small-room acoustics, when large amounts of receiver positions need to be calculated, or when a set of moving receivers needs to be obtained, making the method suitable for scalable multichannel applications.

7. REFERENCES

[1] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*, Springer-Verlag, New York, 2006.

[2] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Acoust. Soc. Amer.*, vol. 53, no. 11, pp. 1004–1025, 2005.

[3] R. Bracewell, *The Fourier Transform & Its Applications*, McGraw-Hill, New York, 3rd edition, 1999.

[4] 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, Jul. 2010.

[5] J. Allen and D. Berkley, "Image method for efficiently simulating small room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1979.

[6] J. Martinez, R. Heusdens, and R. C. Hendriks, "A generalized poisson summation formula and its application to fast linear convolution," *IEEE Signal Process. Lett.*, vol. 18, no. 9, pp. 501 –504, Sept. 2011.

[7] P. Morse and K. Ingard, *Theoretical Acoustics*, McGraw-Hill, New York, 1968.

[8] 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.

[9] P. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, vol. 80, pp. 1527–1529, Nov. 1986.

[10] J. Borish, "Extension of the image method to arbitrary polyhedra," *J. Acoust. Soc. Amer.*, vol. 75, no. 6, pp. 827–1836, 1985.