Audio Engineering Society

Convention Paper 10582

Presented at the 152nd Convention
2022 May, In-Person and Online

Spatial extrapolation of early room impulse responses with source radiation model based on equivalent source method

Izumi Tsunokuni[1], Haruka Matsuhashi[1], and Yusuke Ikeda[1]

[1]*Tokyo Denki University*

Correspondence should be addressed to Izumi Tsunokuni (21udc02@ms.dendai.ac.jp)

ABSTRACT

The measurement of room impulse responses (RIRs) at multiple points is useful in most acoustic applications, such as sound field control. Recently, several methods have been proposed to estimate multiple RIRs. However, when using a small number of closely located microphones, the estimation accuracy degrades owing to the source directivity. In this study, we propose an RIR estimation method using a source radiation model based on the sparse equivalent source method (ESM). First, based on the sparse ESM, the source radiation was modeled in advance by the microphone array enclosing the sound source. Subsequently, the sound field, including the sound reflections, was modeled using the source radiation model based on the sparse ESM and the image source method. As observed from the simulation experiments, the estimation accuracy was improved at higher frequencies compared with the sparse ESM without the source radiation model.

1 Introduction

Room impulse responses (RIRs) at multiple points are used in various acoustic applications, such as sound field reproduction, sound visualization, and room acoustics design. Generally, spatial sound propagation is assumed as a linear time-invariant system. An RIR, or a system response, characterizes a sound propagation between a sound source and a receiver's position. When a measurement signal, such as a sweep-sine signal, is radiated from a loudspeaker, RIR is measured by a microphone at a receiver's position.

When measuring RIR at multiple measurement points for detailed information on the propagation of sound in space, it is required to determine the spacing of the measurement points in consideration of the spatial Nyquist frequency. In the higher frequency, the spacing of the measurement points decreases since the wavelength becomes shorter. Thus, especially when the measurement area expands, the number of microphones and the number of measurements becomes enormous.

Several interpolation methods have been proposed to efficiently obtain RIRs in an area [1, 2, 3, 4]. For example, in [1], the RIRs within an entire volume were interpolated using the sparsity of the early reflections in the time domain. In 2017, Antonello *et al.* proposed an interpolation method using a time-domain equivalent source method (TESM) [2]. Furthermore, an estimation method was proposed based on interpolating the wave field using the prior information of the sound source direction [3, 4].

It is impractical to place numerous microphones in distributed locations to interpolate RIRs. It is desirable to place the microphones in close enough locations to estimate the RIRs around those microphones. However,

the spatial extrapolation of the RIRs is generally more difficult compared to the interpolation. In [5], interpolation and extrapolation methods were proposed to estimate RIRs in a rectangular room using the common-acoustical-pole and residue model. In recent years, with the development of compressive sensing [6], various methods for sound-field estimation using sparsity have been proposed [7, 8, 9, 10, 11, 12, 13]. In particular, Verburg *et al.* focused on the extrapolation method and the reconstruction of the sound field with a superposition of the sparse plane waves [8].

We have proposed an extrapolation method to efficiently measure the multiple early RIRs using a small number of microphones [14]. In our previous work, The early RIRs were modeled using the sparse equivalent source method (ESM) [7] and the image source method [15]. This method improved the extrapolation accuracy of the early RIRs at higher frequencies compared to the conventional extrapolation methods. However, as the frequency increased, the estimation accuracy decreased because of the sound source directivity. Various methods and software tools that provide simulations of RIRs, considering multiple directivity patterns of the sound source, have been proposed [16, 17]. By contrast, the source directivity has not been considered in conventional modeling methods for actually measured RIR.

In this study, we propose an extrapolation method for early RIRs using prior modeling of the source radiation based on the sparse ESM. The objective of this study is to improve the estimation accuracy of the RIRs at higher frequencies and extend the estimation area by reducing the degradation caused by the directivity of the sound source. Simulation experiments were conducted to evaluate the proposed method, considering only direct sound and primary reflections for simplicity

2 Methods

To estimate the RIRs around a microphone array, the sound field is modeled from a set of measured RIRs based on the sparse ESM and the room geometry. The proposed method consists of two steps. The first step consist of modeling the radiation from a sound source including the source directivity. Secondly, the sound field is modeled including the direct sound and sound reflections by using the source radiation model. The details of these two steps are described in the following sections.

2.1 Modeling of a sound source radiation

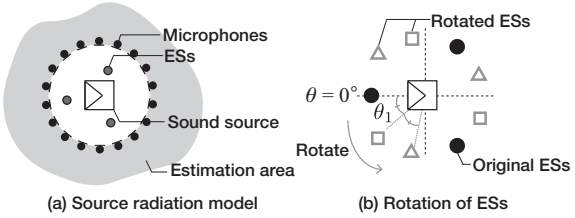

Fig. 1: Modeling of a source radiation. (a) Measurement arrangement. Based on the sparse ESM, the source radiation can be represented through the superposition of the equivalent sources (ESs) around the source position. (b) Use of the source radiation model. In modeling the sound field, the direct sound can be represented using a combination of the rotated ESs of the source radiation model.

Figure 1 (a) shows the concept of modeling the sound source radiation based on the sparse ESM. Here, the measurement is conducted in a non-reverberant room, such as an anechoic chamber. Based on the ESM, the sound field radiating from a sound source, such as a loudspeaker, is represented by a superposition of the equivalent sources. Considering the exterior problem of the Kirchhoff-Helmholtz integral equation, the microphones on the surface enclosing a sound source have sufficient information to model the sound radiation outside the surface [18]. Thus, the M_{src} microphones are placed enclosing the sound source. In addition, N_{src}-equivalent sources are randomly placed around the known position of the sound source.

In frequency domain, the microphone signals at each frequency are represented by the transfer functions of the equivalent sources as follows:

$$
\begin{aligned}
\mathbf{d} &= \mathbf{G}\mathbf{w}, \qquad (1)\\
\mathbf{d} &= [d_1,\ldots,d_m,\ldots,d_{M_{\mathrm{src}}}]^{\mathsf{T}},\\
\mathbf{G} &= \begin{bmatrix} G(\mathbf{x}_{\mathrm{m}}^{(1)},\mathbf{x}_{\mathrm{es}}^{(1)}) & \cdots & G(\mathbf{x}_{\mathrm{m}}^{(1)},\mathbf{x}_{\mathrm{es}}^{(N_{\mathrm{src}})}) \\ \vdots & \ddots & \vdots \\ G(\mathbf{x}_{\mathrm{m}}^{(M_{\mathrm{src}})},\mathbf{x}_{\mathrm{es}}^{(1)}) & \cdots & G(\mathbf{x}_{\mathrm{m}}^{(M_{\mathrm{src}})},\mathbf{x}_{\mathrm{es}}^{(N_{\mathrm{src}})}) \end{bmatrix}\\
\mathbf{w} &= [w_1,\ldots,w_n,\ldots,w_{N_{\mathrm{src}}}]^{\mathsf{T}},
\end{aligned}
$$

where, d_m denotes m-th microphone signal, namely, the complex sound pressure at m-th microphone and $\mathbf{G} = \{G(\mathbf{x}_{\mathrm{m}}^{(m)},\mathbf{x}_{\mathrm{es}}^{(n)})\}(m = 1,\ldots,M_{\mathrm{src}}, n = 1,\ldots,N_{\mathrm{src}})$ is

the matrix of the transfer functions from the equivalent sources to the microphones. $\mathbf{x}_{\mathrm{m}}^{(m)}$ and $\mathbf{x}_{\mathrm{es}}^{(j)}$ are the positions of the m-th microphone and j-th equivalent sources, respectively. w_n is weight coefficient of n-th equivalent source. The transfer functions of the equivalent sources between the equivalent sources and the microphones are calculated using Green's functions in the free field condition. In the three-dimensional sound field, Green's function at any point $\mathbf{x}$ from an equivalent source at $\mathbf{x}_{\mathrm{es}}$ is defined by

$$G(\mathbf{x}, \mathbf{x}_{\mathrm{es}}) = \frac{1}{4\pi} \frac{e^{-jk|\mathbf{x}-\mathbf{x}_{\mathrm{es}}|}}{|\mathbf{x}-\mathbf{x}_{\mathrm{es}}|}, \tag{2}$$

where, j and k denote the imaginary unit and wavenumber, respectively.

Since the number of equivalent sources is significantly larger than that of the real sound source, the vector of the weight coefficients $\mathbf{w}$ is assumed to be sparse. Thus, to obtain the weight coefficients, the following optimization problem is solved.

$$\begin{aligned} &\underset{\mathbf{w}}{\text{minimize}} \quad \|\mathbf{w}\|_1 \\ &\text{subject to} \quad \|\mathbf{d}-\mathbf{G}\mathbf{w}\|_2 \leq \varepsilon_{\mathrm{src}}, \end{aligned} \tag{3}$$

where, $\varepsilon_{\mathrm{src}}$ denotes the error tolerance, and the operators $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the $\ell 1$ and $\ell 2$-norm, respectively. Therefore, the radiation of the sound source can be modeled by the superposition of the equivalent sources with a weight vector $\mathbf{w}$.

2.2 Modeling the sound field

In the next step, to estimate the RIRs around the microphone array, the sound field including direct sound and reflections is modeled from the a small number of RIRs using the source radiation model. We assume that the approximate positions of the sound source, walls, and microphones are known in advance. This assumption is usually reasonable because this approximate geometry can be measured when a microphone array and a sound source are installed.

Figures 1 (b) and 2 show the concept of the proposed method based on the sparse ESM [7]. First, we consider the modeling of the direct sound component. Since the direction of the sound source is not precisely known, the candidates of the equivalent sources are set by rotating the equivalent sources of the radiation model at

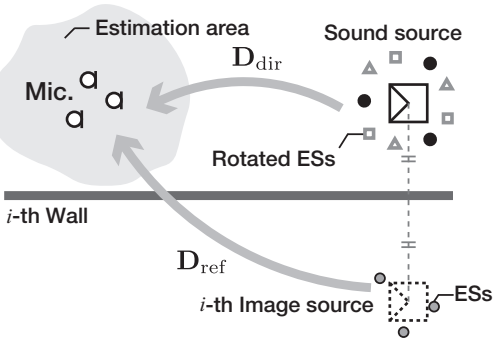

Fig. 2: Concept of the proposed method for modeling the sound field. Based on the sparse ESM, the superposition of the sparse equivalent sources represents the direct and reflected sounds. For the direct sound, the ESs are located by rotating the equivalent sources of the source radiation model, as shown in Fig. 1 (b). For the reflected sounds, the ESs are randomly placed around each image source.

a sufficiently fine angle, as shown in Fig. 1 (b). Thus, the direct sound can be represented as follows.

$$\begin{aligned} &\mathbf{y}_{\mathrm{dir}} = \mathbf{D}_{\mathrm{dir}}\mathbf{u}_{\mathrm{dir}} \\ &\mathbf{D}_{\mathrm{dir}} = [\mathbf{G}^{(\theta_1)}\mathbf{w}, \ldots, \mathbf{G}^{(\theta_\alpha)}\mathbf{w}]\ (\theta = \theta_1, \theta_2, \ldots, \theta_\alpha), \end{aligned} \tag{4}$$

where, $\mathbf{y}_{\mathrm{dir}} (\in \mathbb{C}^{M\times 1})$ denotes the microphone signal vector of direct sound, $\mathbf{G}^{(\theta)} (\in \mathbb{C}^{M\times N_{\mathrm{src}}})$ is the matrix of the transfer functions from θ-rotated equivalent sources to the microphones, and $\mathbf{u}_{\mathrm{dir}} (\in \mathbb{C}^{\alpha\times 1})$ is the weight vector of the rotated equivalent sources. $\theta = 0^\circ$ is the reference of source direction in the first step.

Next, we consider the modeling of the reflected sounds using the sparse ESM [7] and the image-source method [15]. Based on the geometrical acoustics, the positions of the image sources can be calculated to represent the sound reflections. Then, the equivalent sources are randomly positioned around each image source. The sound propagation from all image sources is represented as follows:

$$\mathbf{y}_{\mathrm{ref}} = \mathbf{D}_{\mathrm{ref}}\mathbf{u}_{\mathrm{ref}}, \tag{5}$$

$$\mathbf{D}_{\mathrm{ref}} = [\mathbf{D}_{\mathrm{ref}}^{(1)}, \mathbf{D}_{\mathrm{ref}}^{(2)}, \ldots, \mathbf{D}_{\mathrm{ref}}^{(I)}]\ \ (i = 1, 2, \ldots, I) \tag{6}$$

where, $\mathbf{y}_{\mathrm{ref}} (\in \mathbb{C}^{M\times 1})$ denotes the microphone signal vector of the reflected sounds, $\mathbf{D}_{\mathrm{ref}}^{(i)} = G(\mathbf{x}_{\mathrm{m}}^{(m)}, \mathbf{x}_{\mathrm{es}}^{(n)})(m = 1, \ldots, M, n = 1, \ldots, N)$ is the matrix of the transfer functions between the equivalent sources positioned

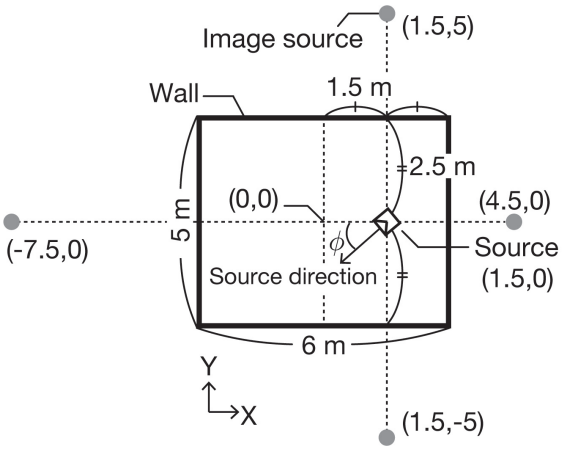

Fig. 3: Geometry of room and image sources in the simulation experiment. The sound source is positioned at $(1.5,0)$. Image sources for primary reflections are considered in the X-Y plane. Equivalent sources are positioned randomly around each image source.

around i-th image source and the microphones, and $\mathbf{u}_{\mathrm{ref}}(\in \mathbb{C}^{(I\times N)\times 1})$ is the weight vector of equivalent sources. N is the number of equivalent sources for each image source and I indicates the number of image sources.

Based on the linearity of the sound, the measured microphone signals $\mathbf{y}$ are a combination of the direct and reflected sounds: $\mathbf{y} = \mathbf{y}_{\mathrm{dir}} + \mathbf{y}_{\mathrm{ref}}$. Thus, from Eq. (4) and (5), the following equation is derived:

$$\mathbf{y} = \mathbf{D}\mathbf{u}, \tag{7}$$

$$\mathbf{D} = [\mathbf{D}_{\mathrm{dir}}, \mathbf{D}_{\mathrm{ref}}] \tag{8}$$

$$\mathbf{u} = [\mathbf{u}_{\mathrm{dir}}^{\mathsf{T}}, \mathbf{u}_{\mathrm{ref}}^{\mathsf{T}}]^{\mathsf{T}} \tag{9}$$

The number of columns for the matrix $\mathbf{D}$ is the total number of equivalent sources for the direct and reflected sounds, $N' = (\alpha + (I \times N))$. The weight coefficients $\mathbf{u}$ are assumed to be sparse. This is because the single sound source is directed in one direction, and the image sources are sparsely positioned. Thus, to solve the weights of all the equivalent sources that are positioned randomly around the source and image sources, the optimization problem is formulated as follows:

$$\begin{aligned} \underset{\mathbf{v}}{\text{minimize}} \quad & \|\mathbf{u}\|_1 \\ \text{subject to} \quad & \|\mathbf{y} - \mathbf{D}\mathbf{u}\|_2 \leq \varepsilon_{\mathrm{sf}}, \end{aligned} \tag{10}$$

where, $\varepsilon_{\mathrm{sf}}$ is the error tolerance. The estimated early RIR $\hat{y}$ at any point $\mathbf{x}$ is estimated by using all equivalent sources as follows:

$$\hat{y}(\mathbf{x}) = \sum_{n=1}^{N'} u_n G(\mathbf{x}, \mathbf{x}_{\mathrm{es}}^{(n)}), \tag{11}$$

where u_n, which is the n-th element of the vector $\mathbf{u}$, denotes the weight coefficient of the n-th equivalent source positioned at $\mathbf{x}_{\mathrm{es}}^{(n)}$.

3 Simulation experiment

3.1 Experimental conditions

The simulation experiments were conducted using the image source method to evaluate the estimation accuracy of the proposed method. We consider only the direct sound and the primary reflections in a two-dimensional sound field for simplicity. Note that a sound propagates in a three-dimension as a point source, and the evaluated area is in X-Y plane. The proposed method was compared with the conventional method without the source radiation model [14] and the plane wave approximation [8].

Table 1 and Fig. 3 show the simulation conditions. The center of the microphone array is the origin of the coordinates. The room geometry was rectangular, with a size of 5 m $\times$ 6 m. The source was positioned at $(X,Y) = (1.5,0)$ m. We simulated up to the primary reflection, thus, four image sources were considered, as shown in Fig. 3. In this simulation, the source directivity was unidirectional for simplicity. Thus, the amplitude of the source was changed by the direction of the microphone position. When the microphone's direction is ϕ, the amplitude of the sound source is represented by $\cos(1+\phi)/2$, where, the origin direction is $\phi = 0$ [rad]. Nine hundred equivalent sources were positioned around the sound source and each image source. The equivalent sources for the direct sound, which were obtained in the first step of the proposed method, were rotated by 10 degree in the range of $\pm 30^{\circ}$.

The noises with an amplitude SNR of 30 dB and uniform random phases were added to the microphone signals in the sound field model. Twenty simulations were conducted and the results were averaged. The estimation errors are defined by,

$$\mathrm{Error}(\mathbf{x},\omega) = 10\log_{10}\frac{|\hat{p}(\mathbf{x},\omega) - p(\mathbf{x},\omega)|^2}{|p(\mathbf{x},\omega)|^2}, \tag{12}$$

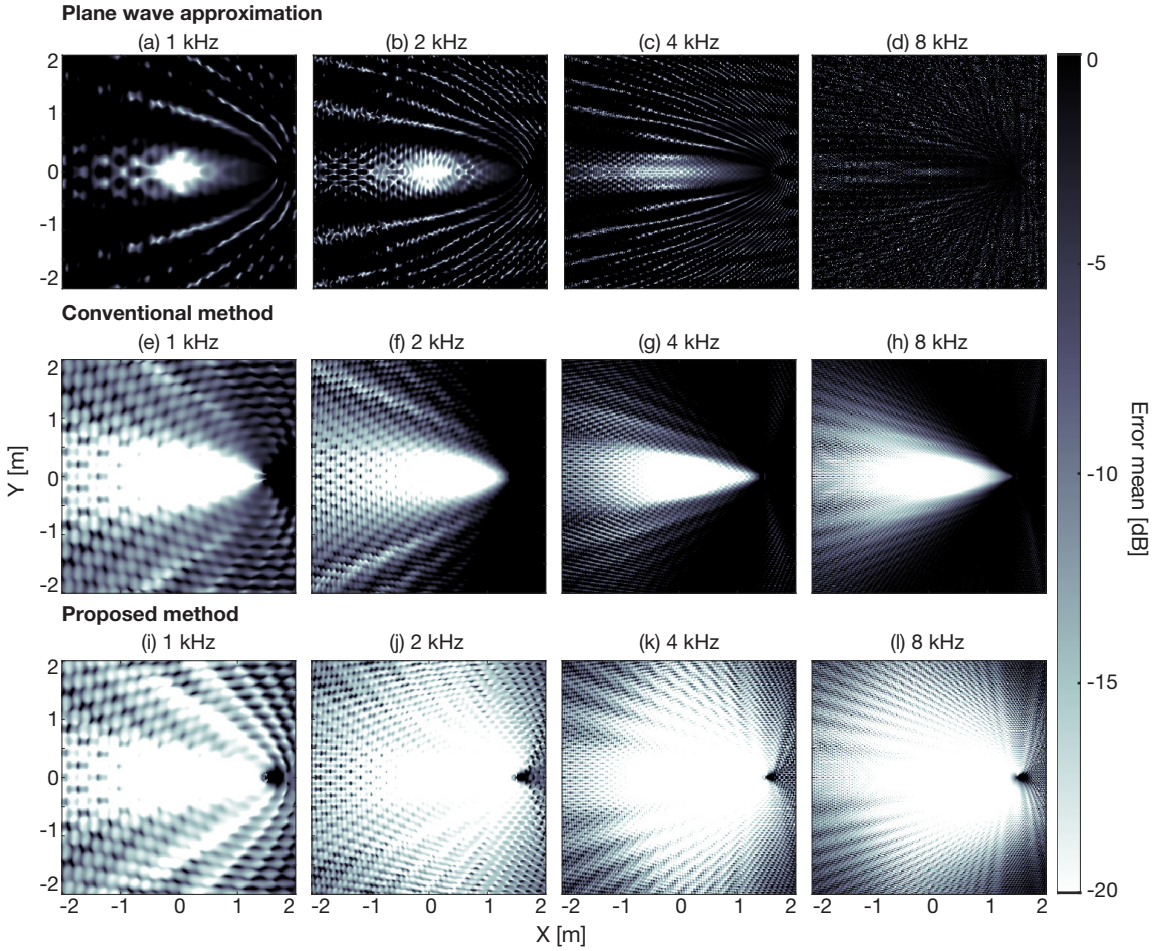

Fig. 4: Distributions of the estimation errors at 1, 2, 4 and 8 kHz. The source is positioned at $(1.5, 0)$. The cross-shaped microphone array is centered in the estimated region. The upper figures indicate the results of the plane-wave approximation. The middle figures indicate the results of the sparse ESM with no source radiation model (conventional method). The bottom figures show the results of the proposed method using the source radiation model.

where, p and $\hat{p}$ denote the desired and estimated complex sound pressures at $\mathbf{x}$ at the frequency ω, respectively. The error tolerances in Eq. (3) and (10) were decided by the $\ell 2$-norm of the microphone signal's noise. The simulations were performed using MATLAB (R2021a) with CVX (ver. 2.2) based on the interior point method, which is a package for specifying and solving convex programs [19, 20].

3.2 Experimental results

First, the error distributions in the evaluation area were compared. Figure 4 shows the distributions of estimation errors in 4 m $\times$ 4 m area at 1, 2, 4 and 8 kHz. As shown in the upper figures of Fig. 4, the error of the plane wave approximation was less than -20 dB only near the microphones at 1 and 2 kHz. However, above 4 kHz, the errors over the evaluation area is significantly degraded. Thus, plane wave approximation can estimate the RIRs only near the microphones at low frequencies.

In the conventional method and the proposed method, the errors of both the methods were less than -20 dB over a broader range than that of the plane wave approximation at each frequency. In the conventional method, the estimable area spreads radially from the source because of the source directivity and becomes

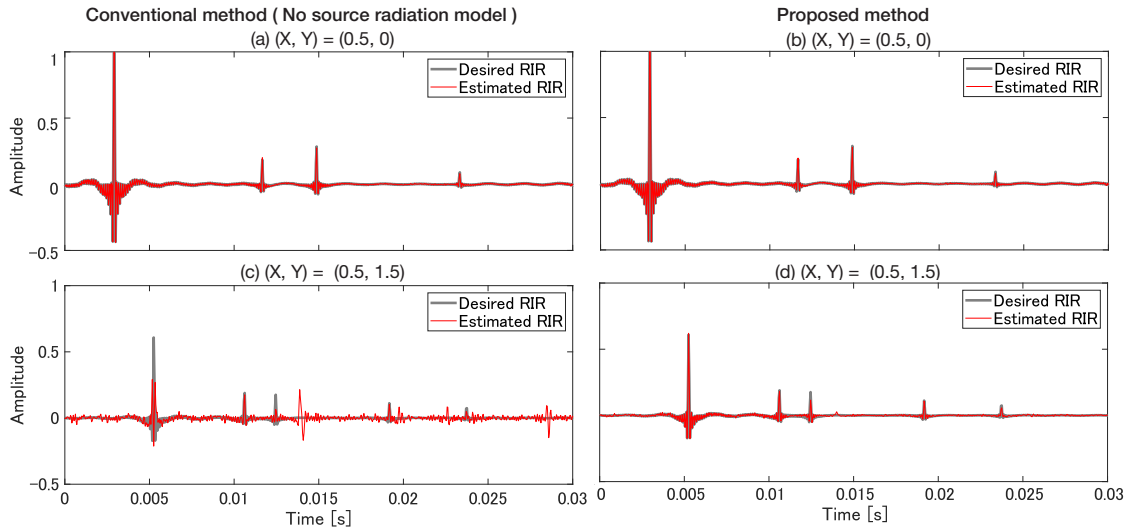

Fig. 5: Comparison of room impulse responses estimated by the conventional and proposed method. (a) and (b) show the RIRs at (0.5, 0) near the microphones. (c) and (d) show the RIRs at (0.5, 1.5), namely, farther away from the microphones.

Table 1: Simulation conditions

	Modeling	
	Source Radiation	Sound Field
Microphone array	Circle (Radius 0.5 m)	Cross (One side 0.6 m)
Number of microphones	$M_{\mathrm{src}} = 36$	$M = 25$
Source direction	$\phi = 0°$	$\phi = 5°$
Noise in Mic. signal	Amplitude SNR 50 dB Phase random	Amplitude SNR 30 dB Phase random

rapidly narrower above 2 kHz. In contrast, in Fig. 4 (i), (j), (k), and (l), the estimable area in the proposed method was spread out centering on the microphone array. Comparing the two methods, the improvements in the proposed method were noticeable above 2 kHz.

Notably, the errors at 1 kHz were larger than those above 2 kHz in the proposed method. This is probably caused by the relationship between the array size and the wavelength at 1 kHz. The array size was 0.6 m, which was less than twice the wavelength of 1 kHz. Thus, the estimation was sensitive to noise because the obtained spatial information was not sufficiently broad for the wavelength.

Here, the RIR signals in the time domain are compared using the conventional and proposed methods. Figure 5 shows the estimated and desired RIR signals near and away from microphones. Comparing Fig. 5 (a) and (b), RIRs near the microphones were accurately estimated in the conventional and proposed method. The estimated RIR signal and the desired signal were consistent in amplitude and time of arrival for direct and reflected sounds.

Fig. 5 (c) and (d) show the estimated and desired RIRs away from microphones. The amplitude of direct sound arrived at approximately 5 ms was half energy of the desired signal, as shown in Fig. 5(c). The reflected sound at approximately 15 ms also deviated from the desired one in amplitude and arrival time. In addition, the artifacts were observed in the estimated RIR as some pulses, for example, a pulse after 25 ms. By contrast, the pulse signal was accurately estimated for both the direct and reflected sounds in the proposed method, as shown in Fig. (d).

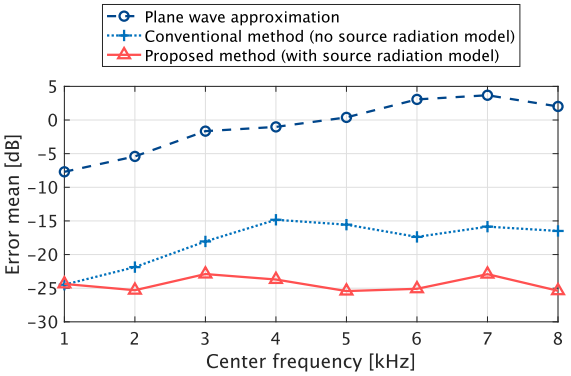

Fig. 6: The mean errors in each 1 kHz band at 1m × 1m around microphones. The number of evaluation points is 5,625. The interval of evaluation points is approximately 0.01m.

Figure 6 shows the error means of the 1 kHz frequency bands. The conventional method (no source radiation model) and the proposed method had smaller errors compared to the plane wave approximation over all the frequency bands. At the 1 kHz-centered band, there were no significant large differences in the errors between the conventional method and the proposed method. However, the errors in the conventional method increased above the 2 kHz band. As the frequency increased up to the 4 kHz band, the errors in the conventional method increased. In contrast, the estimation accuracies of the proposed method were nearly constant over all the bands. At the 8 kHz-centered band, the estimation accuracy of the proposed method improved by approximately 7 dB and 25 dB, respectively, compared to the conventional method and the plane wave approximation. Therefore, the proposed method improved the estimation accuracies of the sound field with a directional sound source, especially at frequencies above 2 kHz.

4 Summary and future works

In this study, we proposed an estimation method for early RIRs around a microphone array by modeling RIRs with a prior source radiation model. The proposed method can estimate the early RIRs using a combination of the sparse equivalent sources and the image source methods. The simulation experiments showed that by modeling the source radiation characteristics in advance, the estimation accuracies were improved by approximately 7 dB at the 4–8 kHz-centered frequency bands compared to the method with the no source radiation model. In addition, the estimable region of higher accuracy was larger at frequencies above 2 kHz. In future work, we will extend the proposed method to higher-order reflections with an evaluation using the measurement of the RIRs in an actual environment.

Acknowledgement

This work was partially supported by JSPS KAKENHI Grant Number 19K12049 and 22K12099.

References

[1] Mignot, R., Daudet, L., and Ollivier, F., “Room Reverberation Reconstruction: Interpolation of the Early Part Using Compressed Sensing,” *IEEE Transactions on Audio, Speech, and Language Processing*, 21(11), pp. 2301–2312, 2013.

[2] Antonello, N., De Sena, E., Moonen, M., Naylor, P. A., and van Waterschoot, T., “Room Impulse Response Interpolation Using a Sparse Spatio-Temporal Representation of the Sound Field,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(10), pp. 1929–1941, 2017.

[3] Ito, H., Koyama, S., Ueno, N., and Saruwatari, H., “Spatial Active Noise Control Based on Kernel Interpolation with Directional Weighting,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8404–8408, 2020.

[4] Ueno, N., Koyama, S., and Saruwatari, H., “Directionally Weighted Wave Field Estimation Exploiting Prior Information on Source Direction,” *IEEE Transactions on Signal Processing*, 69, pp. 2383–2395, 2021.

[5] Haneda, Y., Kaneda, Y., and Kitawaki, N., “Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function,” *IEEE Transactions on Speech and Audio Processing*, 7(6), pp. 709–717, 1999.

[6] Candes, E. J. and Wakin, M. B., “An Introduction To Compressive Sampling,” *IEEE Signal Processing Magazine*, 25(2), pp. 21–30, 2008.

[7] Fernandez-Grande, E., Xenaki, A., and Gerstoft, P., "A sparse equivalent source method for near-field acoustic holography," *The Journal of the Acoustical Society of America*, 141(1), pp. 532–542, 2017.

[8] Verburg, S. A. and Fernandez-Grande, E., "Reconstruction of the sound field in a room using compressive sensing," *The Journal of the Acoustical Society of America*, 143(6), pp. 3770–3779, 2018.

[9] Katzberg, F., Mazur, R., Maass, M., Koch, P., and Mertins, A., "A Compressed Sensing Framework for Dynamic Sound-Field Measurements," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(11), pp. 1962–1975, 2018.

[10] Koyama, S. and Daudet, L., "Sparse Representation of a Spatial Sound Field in a Reverberant Environment," *IEEE Journal of Selected Topics in Signal Processing*, 13(1), pp. 172–184, 2019.

[11] Zea, E., "Compressed sensing of impulse responses in rooms of unknown properties and contents," *Journal of Sound and Vibration*, 459, p. 114871, 2019.

[12] Mazur, R., Katzberg, F., and Mertins, A., "Robust Room Equalization Using Sparse Sound-field Reconstruction," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4230–4234, 2019.

[13] Vinceslas, L., Lim, H., and Kondoz, A., "Evaluation of sparse sound field models for compressed sensing in multiple sound zones," in *Audio Engineering Society Convention 148*, 2020.

[14] Tsunokuni, I., Kurokawa, K., Matsuhashi, H., Ikeda, Y., and Osaka, N., "Spatial extrapolation of early room impulse responses in local area using sparse equivalent sources and image source method," *Applied Acoustics*, 179, p. 108027, 2021.

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

[16] Habets, E. A. P., "Room Impulse Response Generator," AudioLabs, Erlangen, Tech. Rep., 2010.

[17] Savioja, L. and Svensson, U. P., "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, 138(2), pp. 708–730, 2015.

[18] Williams, E. G., *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Academic Press, 1999, ISBN 0127539603.

[19] Grant, M. and Boyd, S., "CVX: Matlab Software for Disciplined Convex Programming, version 2.2," `http://cvxr.com/cvx`, 2020.

[20] Grant, M. and Boyd, S., "Graph implementations for nonsmooth convex programs," in V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pp. 95–110, Springer-Verlag Limited, 2008, `http://stanford.edu/~boyd/graph_dcp.html`.