Room Impulse Response Interpolation from a Sparse Set of Measurements Using a Modal Architecture

Orchisama Das [1] Paul Calamia [2] Sebastia V. Amengual Gari [2]

[1]Center for Computer Research in Music and Acoustics, Stanford University, Stanford, USA

[2]Facebook Reality Labs Research Audio, Redmond, USA

IEEE International Conference on Acoustics, Speech and Signal Processing

Toronto, Canada, June 6-11, 2021

Objectives

- Characterizing the entire sound field of a room utilizing a sparse dataset of Room Impulse Responses (RIRs) measured at different locations.
- Accurate spatial interpolation of perceptually relevant low frequency modes in rooms with simple geometries having non-rigid walls.
- Useful for real-time interpolation and extrapolation in Augmented Reality (AR) applications.

Room Modes

- Solutions to the 3D wave equation are standing waves, or room modes.
- The RIR can be characterized by a sum over M modes, whose complex amplitudes, γ, are functions of space, whereas frequencies and dampings, ω and α, determine the temporal response:

$$h(x,y,z,t) = \sum_{m=1}^{M} \gamma_m(x,y,z) \exp\left[(j\omega_m - \alpha_m)t\right].$$

- Complex mode amplitudes are the solution to the homogeneous Helmholtz equation [1]:

$$\gamma_m(\mu) = C_{\mu_m} \exp(jk_{\mu_m}\mu) + D_{\mu_m} \exp(-jk_{\mu_m}\mu);$$
$$k_{\mu_m} \rightarrow \text{Wave number};\ C_{\mu_m}, D_{\mu_m} \rightarrow \text{Constants};\ \mu \in (x,y,z).$$

- We want to estimate the unknown wave numbers and constants for each mode from a set of RIR measurements at different locations in the room.

Mode Estimation

- RIRs measured at different positions are time-aligned and averaged.
- Common poles (mode frequencies and amplitudes) calculated from the averaged RIR with subband ESPRIT [2].
- Mode amplitudes estimated with linear least squares.

Non-linear Optimization

- For 2D interpolation with measurements at L locations,

$$\begin{bmatrix} \hat{\gamma}_m(x_1,y_1) \\ \hat{\gamma}_m(x_2,y_2) \\ \vdots \\ \hat{\gamma}_m(x_L,y_L) \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_{m_1} \\ \boldsymbol{u}_{m_2} \\ \vdots \\ \boldsymbol{u}_{m_L} \end{bmatrix} \begin{bmatrix} C_{1m} \\ D_{1m} \\ C_{2m} \\ D_{2m} \end{bmatrix}$$

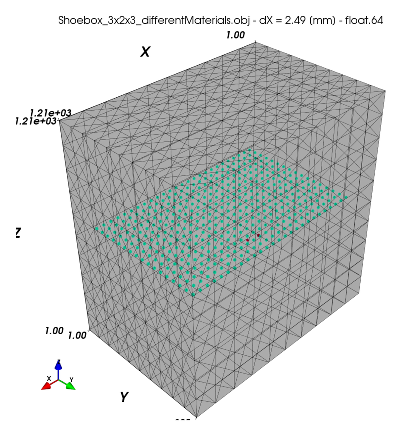
$$\boldsymbol{u}_{m_l} = [e^{-j(k_{x_m}x_l + k_{y_m}y_l)}\ e^{j(k_{x_m}x_l + k_{y_m}y_l)}\ e^{-j(k_{x_m}x_l - k_{y_m}y_l)}\ e^{j(k_{x_m}x_l - k_{y_m}y_l)}]$$
$$\hat{\boldsymbol{\gamma}}_m = \boldsymbol{U}_m(k_{x_m}, k_{y_m})\boldsymbol{c}_m.$$

- Find optimal parameters for the first M_c modes with sequential optimization [3].

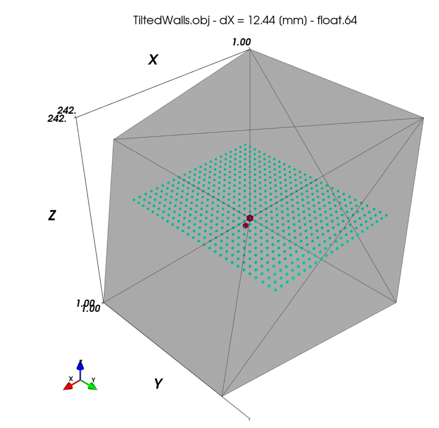
Algorithm 1 Sequential optimization

Require: $0 \leq k_{x_m}, k_{y_m} \leq \frac{\omega_m + j\alpha_m}{c}\ \forall\ m$
for $m = 1 \cdots M_c$ **do**
Initialize $k_{x_m} = k_{y_m} = \frac{\omega_m + j\alpha_m}{\sqrt{2}c}$
repeat
$\boldsymbol{c}_{m_i} = \boldsymbol{U}^{*\dagger}_{m_{i-1}} \boldsymbol{\gamma}_m$
$\hat{\boldsymbol{\gamma}}_{m_i} = \boldsymbol{U}^{*}_{m_{i-1}} \boldsymbol{c}_{m_i}$
$J(k_{x_{m_i}}, k_{y_{m_i}}) = ||20\log_{10}(\boldsymbol{\gamma}_m \oslash \hat{\boldsymbol{\gamma}}_{m_i})||_2^2$
$k^*_{x_{m_i}}, k^*_{y_{m_i}} = \arg\min_{k_{x_m}, k_{y_m}} J$
until convergence
end for

FDTD Simulations



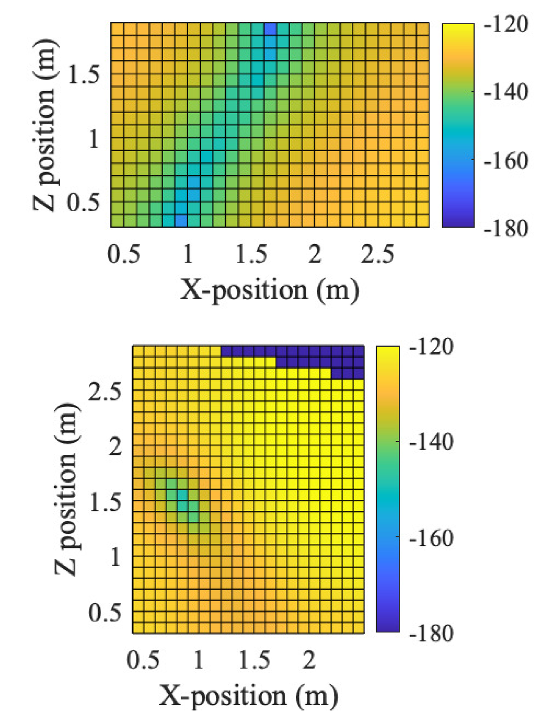
Shoebox room with different materials



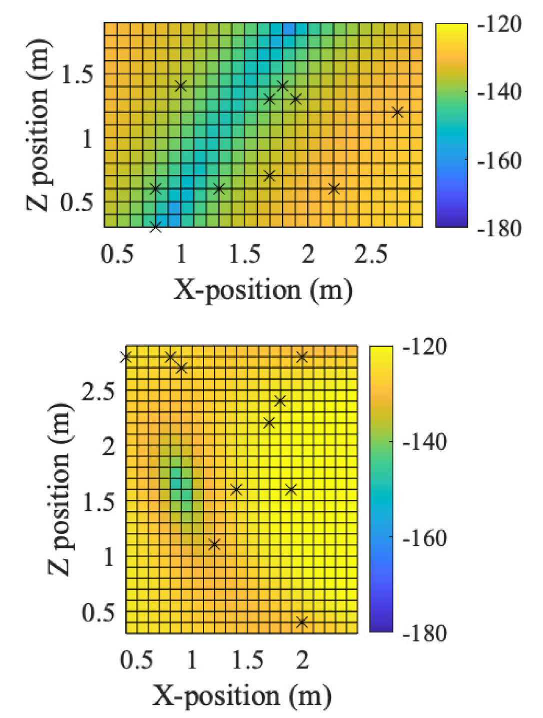
Room with tilted walls.

- Shoebox room of dimensions $3 \times 2 \times 3$ m^3 with different admittances (K_d) on the walls - front and back wall $K_d = 0.9$, left and right wall $K_d = 0.8$, floor and ceiling $K_d = 0.99$.
- Non-rectangular room with no parallel walls, 3 m on the longest edge in each direction, made of the same materials and having tilted walls at an angle of 9.4° in the x, z directions.
- Omni-directional point source and virtual microphones placed in a rectangular grid on the xy plane at a height of 1.7 m.
- Microphone grid resolution is $d = 0.2$ m. Maximum mode frequency corresponding to M_c is $\frac{c}{2d} = 866$ Hz.

Results

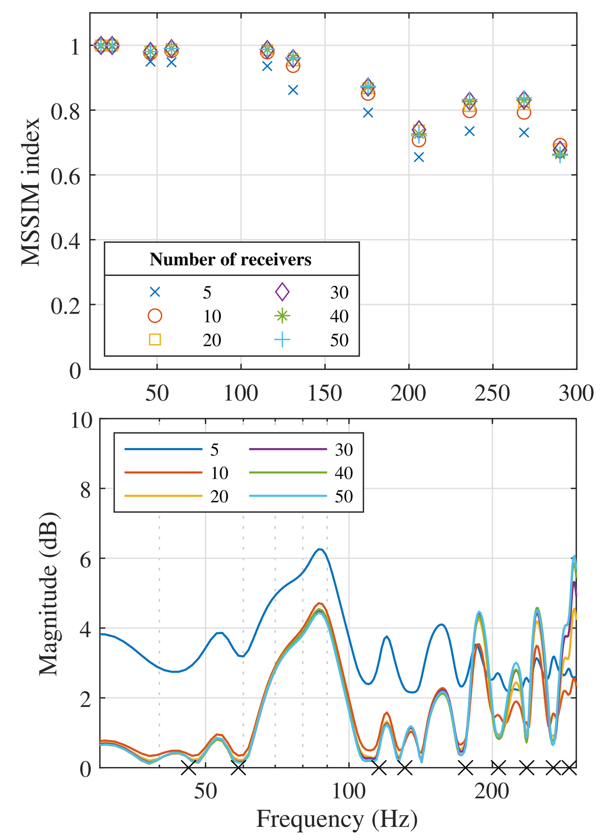


Mode at 58 Hz and 62 Hz, measured.

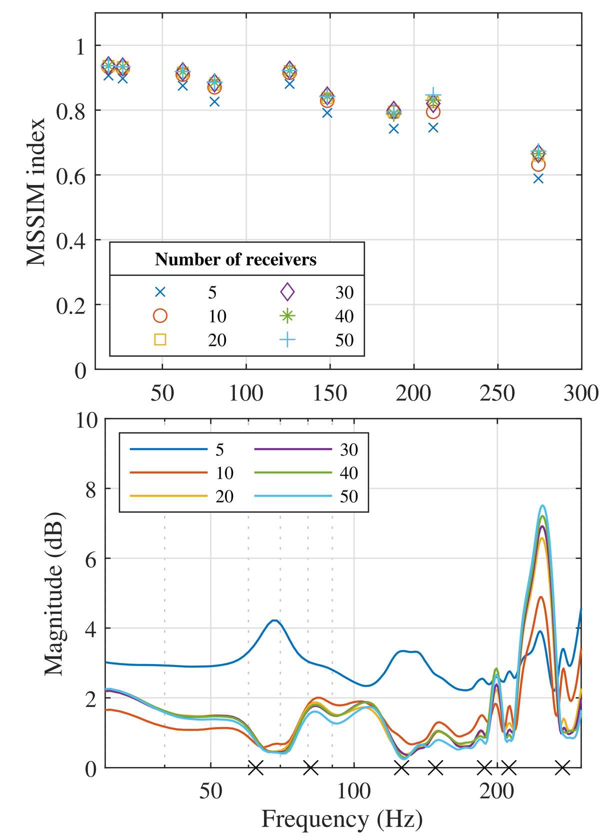


Mode at 58 Hz and 62 Hz, fit.

- Number of microphones varied from 5 to 50. 100 trials ran for each set, microphones placed in different randomized configurations in each trial.
- MSSIM - Mean structural similarity between measured and fit mode shapes.
- AMSDE - Absolute mean spectral difference error. Absolute difference between the frequency responses (averaged over all measurement points and all configurations) of the measured and modeled RIR, expressed in dB.



Shoebox - MSSIM(T), AMSDE(B)



TiltedWalls - MSSIM(T), AMSDE(B)

References

[1] Y. Haneda, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function," *IEEE Trans. on Speech and Audio Process.*, vol. 7, no. 6, pp. 709–717, 1999.

[2] C. Kereliuk, W. Herman, R. Wedelich, and D. J. Gillespie, "Modal analysis of room impulse responses using subband ESPRIT," in *Proceedings of the Int. Conf. on Digital Audio Effects*, 2018.

[3] P. Ainsleigh and J. George, "Modeling exponential signals in a dispersive multipath environment," in *Proceedings of IEEE Int. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, vol. 5. IEEE, 1992, pp. 457–460.