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

Orchisama Das, Paul Calamia, Sebastià V. A. Gari

Motivation

- Dynamic reverberation with changing source and listener position is a requirement for realistic room acoustics simulation.

  - VR

    Information about geometry, materials is available.
  - AR

    Missing geometries, inaccurate materials.

    Sparse representation of the room can be obtained.

Objectives

- Generate a continuous RIR space utilizing a sparse dataset of RIRs from a room.
- Smooth, real-time interpolation, extrapolation and morphing between measured positions.
- Accurate spatial interpolation of perceptually relevant low frequency modes in rooms with simple geometries having non-rigid walls.

Related Work

- RIR interpolation with Dynamic Time Warping.[1]
  - Not a function of source/listener position.

- Compressed Sensing[2]
  - Reconstructs early part of RIR only.
  - Requires accurate calibration of microphone array.

- Common acoustical-pole and residue model[3] (derived from room modes).
  - Parameterized by source and listener locations.
  - Physics is well known for rectangular rooms.

[1] Masterson et al. "Acoustic impulse response interpolation for multichannel systems using dynamic time warping.”, *AES 2009*
[2] Mignot et al. "Room reverberation reconstruction: Interpolation of the early part using compressed sensing." *IEEE TALSP 2013.*
[3] Haneda et al. "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function." *IEEE TASLP, 99*

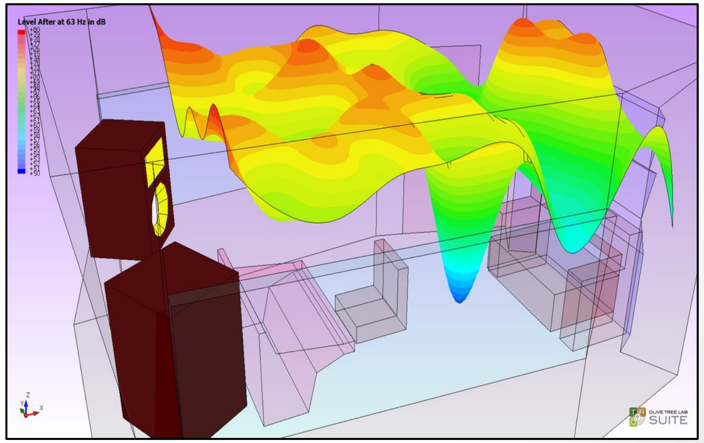
Room Modes I

- Standing waves are solutions to the 3D wave equation
- RIR can be characterized as a sum of M room modes[4]

$$h(x,y,z,t) = \sum_{m=1}^{M} \gamma_m(x,y,z) e^{(j\omega_m - \alpha_m)t}$$

Complex mode amplitudes depend on spatial location of source/listener

Mode frequencies and dampings determine the time response



- RIRs can be efficiently synthesized with M parallel biquads

$$H(z) = \sum_{m=1}^{M} \frac{Re(\gamma_m) - e^{-\alpha_m} Re\left(\gamma_m e^{-j\omega_m}\right) z^{-1}}{1 - 2e^{-\alpha_m}\cos(\omega_m) z^{-1} + e^{-2\alpha_m} z^{-2}}$$

[4] Abel et al. , “A modal architecture for reverberation with application to room acoustics modeling”, *AES 2014*.

Room Modes II

- Complex mode amplitudes are the solution to Helmholtz equation

Wave number

$$\nabla^2 p + k^2 p = 0; \; k = \frac{\omega_m}{c}$$

- For rigid walls

$$p(x,y,z) = C \, \cos(k_x x)\cos(k_y y)\cos(k_z z)$$

From boundary conditions

$$k_\mu = \frac{n_\mu \pi}{L_\mu}$$

$$k = \sqrt{k_x^2 + k_y^2 + k_z^2}$$

- For non-rigid walls with finite impedance

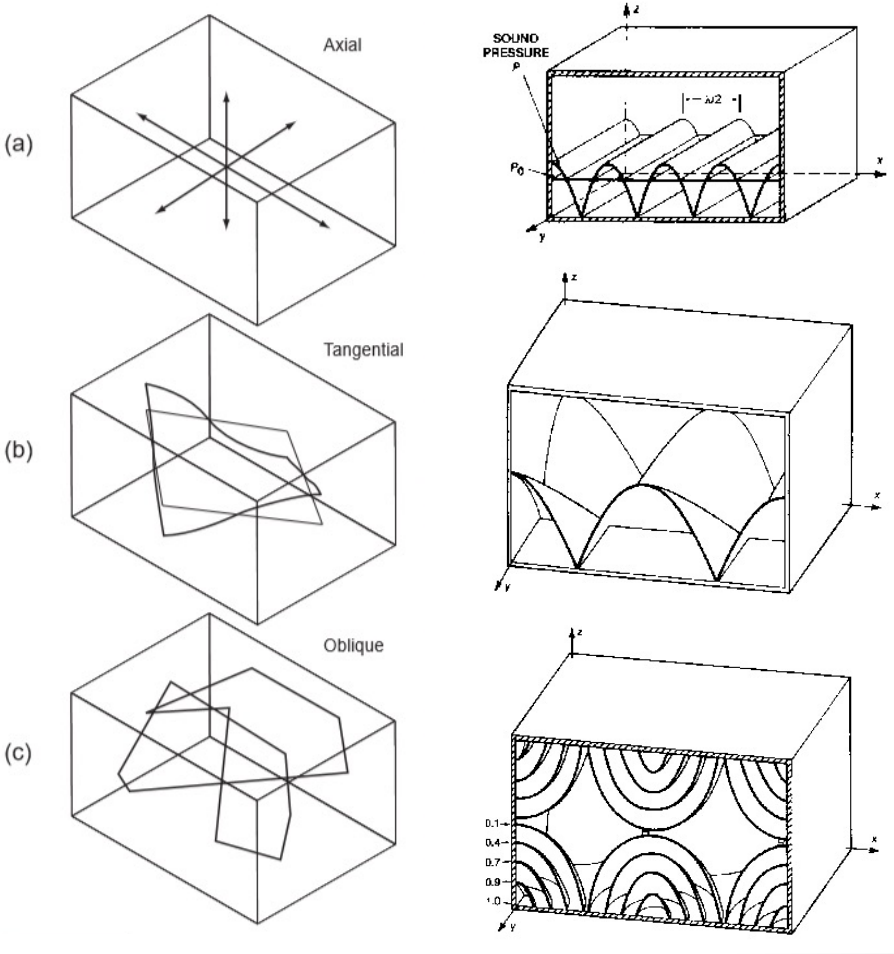
$$p(\mu) = C_\mu \exp\left(i k_\mu \mu\right) + D_\mu \exp\left(-i k_\mu \mu\right)$$

$$k_\mu = \frac{n_\mu \pi}{L_\mu} + j\delta_u$$

Wall absorption

$$p(x,y,z) = p(x)p(y)p(z)$$

Pressure distributions



(4,0,0)

(2,1,0)

(2,1,1)

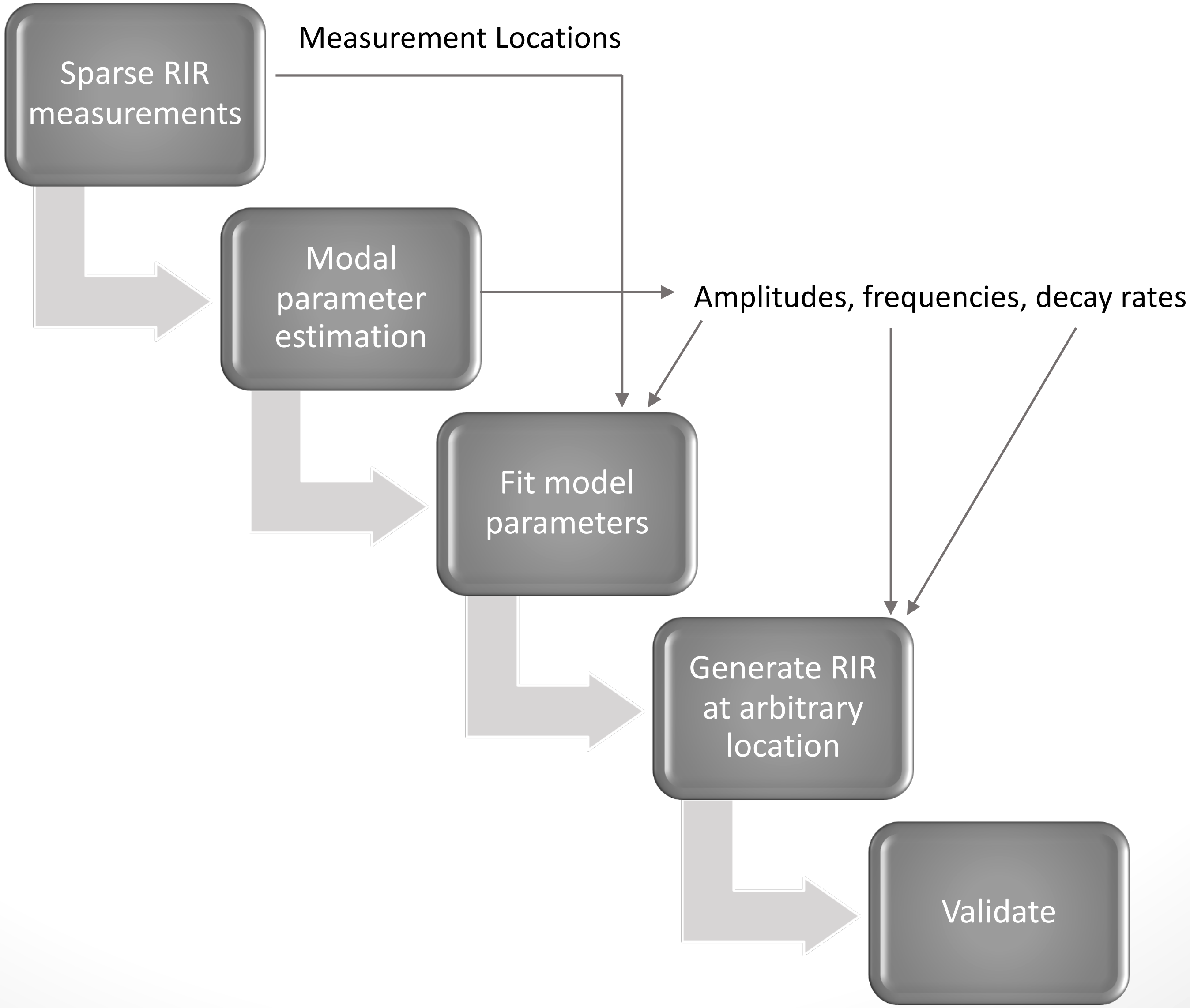
Kuttruff, Heinrich. *Room acoustics, 4th Edition,* Spon Press (2000).

Room Modes III

- Analytic solution for pressure distribution as a function of spatial location in a room (with non-rigid walls) requires
  - knowledge of room dimensions.
  - wall absorption, which is frequency dependent.

- Assuming we don't have this information, can we fit parameters to the measurements?

Pipeline



Sparse RIR measurements
Measurement Locations
Modal parameter estimation
Amplitudes, frequencies, decay rates
Fit model parameters
Generate RIR at arbitrary location
Validate

Modal Estimation

For a grid of RIR measurements

- Average RIRs over all locations and estimate common mode frequencies ω_m and decay rates α_m using subband ESPRIT[5]
- Calculate unique set of M mode amplitudes for L locations in grid using linear least squares.

$$\begin{pmatrix} h_1(0) & h_2(0) & \cdots & h_L(0) \\ h_1(1) & h_2(1) & \cdots & h_L(1) \\ \vdots & \vdots & \vdots & \vdots \\ h_1(T) & h_2(T) & \cdots & h_L(T) \end{pmatrix} = \begin{pmatrix} 1 & \dots & 1 \\ e^{-(j\omega_1-\alpha_1)1} & \cdots & e^{-(j\omega_M-\alpha_M)1} \\ \vdots & \ddots & \vdots \\ e^{-(j\omega_1-\alpha_1)T} & \cdots & e^{-(j\omega_M-\alpha_M)T} \end{pmatrix} \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{L1} \\ \gamma_{12} & \cdots & \gamma_{L2} \\ \vdots & \vdots & \vdots \\ \gamma_{M1} & \cdots & \gamma_{ML} \end{pmatrix}$$

[5] Kereliuk et al."Modal analysis of room impulse responses using subband ESPRIT." In *Proceedings of the International Conference on Digital Audio Effects*. 2018.

Optimization - Model

- Goal - Given a sparse set of impulse responses on a 2D plane, find the complex mode amplitudes and parameterize them as functions of (x,y) location

- General solution to 2D Helmholtz equation for non-rigid walls is

$$\widehat{\gamma_m}(x,y) = C_{1m}\mathrm{e}^{-j(k_{x_m}x + k_{y_m}y)} + D_{1m}\,\mathrm{e}^{j(k_{x_m}x + k_{y_m}y)} + C_{2m}\,\mathrm{e}^{-j(k_{x_m}x - k_{y_m}y)} + D_{2m}\,\mathrm{e}^{j(k_{x_m}x - k_{y_m}y)}$$

- For each location in grid $(x_1, y_1), \cdots, (x_L, y_L)$

$$\begin{pmatrix} \widehat{\gamma_m}\,(x_1, y_1) \\ \vdots \\ \widehat{\gamma_m}\,(x_L, y_L) \end{pmatrix} = \begin{pmatrix} e^{-j(k_{x_m}x_1 + k_{y_m}y_1)} & e^{j(k_{x_m}x_1 + k_{y_m}y_1)} & e^{-j(k_{x_m}x_1 - k_{y_m}y_1)} & e^{-j(k_{x_m}x_1 - k_{y_m}y_1)} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-j(k_{x_m}x_L + k_{y_m}y_L)} & e^{j(k_{x_m}x_L + k_{y_m}y_L)} & e^{-j(k_{x_m}x_L - k_{y_m}y_L)} & e^{-j(k_{x_m}x_L + k_{y_m}y_L)} \end{pmatrix} \begin{pmatrix} C_{1m} \\ D_{1m} \\ C_{2m} \\ D_{2m} \end{pmatrix}$$

$$\widehat{\gamma_m} = U_m\,(k_{x_m}, k_{y_m})\,c_m$$

Linear in constants

Non-linear in wave numbers

Optimization –Sequential update

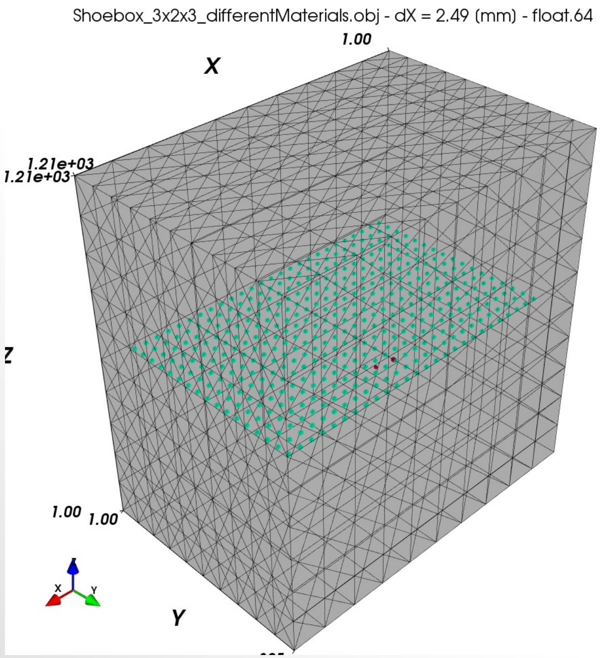
- Find wave numbers and constants for each mode from the measured set of mode amplitudes.

- For modes, m = 1, 2,…, M_c,
  - Initialize : $k_{x_m}(0), k_{y_m}(0) = \frac{\omega_m + j\alpha_m}{\sqrt{2}c}$
  - Repeat, till convergence
    - Update constants : $c_m\,(i) = U_m(i-1)^+\,\gamma_m$
    - Update wave numbers : $\operatorname{argmin}_{k_{x_m}, k_{y_m}} \left|\left|20\log_{10}(\gamma_m) - 20\log_{10}(U_m(i)c_m(i))\right|\right|^2$
    - Bounds : $0 \le k_{x_m}, k_{y_m} \le \frac{\omega_m + j\alpha_m}{c}$

- Calculate mode amplitudes at arbitrary 2D location in room
  - $\widehat{\gamma_m}(x, y) = U_m\left(k_{x_m}^*, k_{y_m}^*\right) c_m^*$

Validation

- Accurate simulation of small room acoustics with FDTD simulator.[6]
- Omni source placed at (0.3,0.2,0.1). Receivers placed in XY plane at resolution of *d* = 20cm at a nominal standing height of 1.7m.
- Maximum mode frequency corresponding to M_c is $\frac{c}{2d} = 866$ Hz.



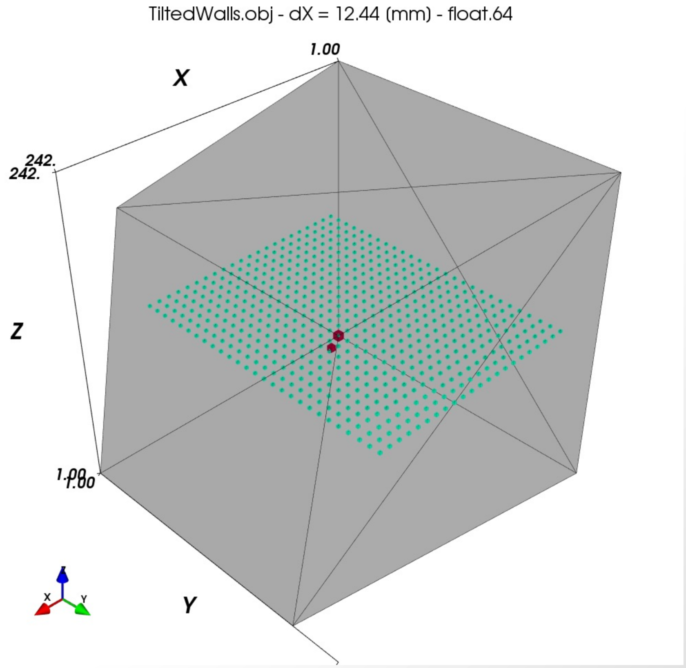
Left – Shoebox room of dimensions 3x2x3 m^3

Right – Room with no parallel walls, 3m on longest edge

Front and back wall admittance - 0.9

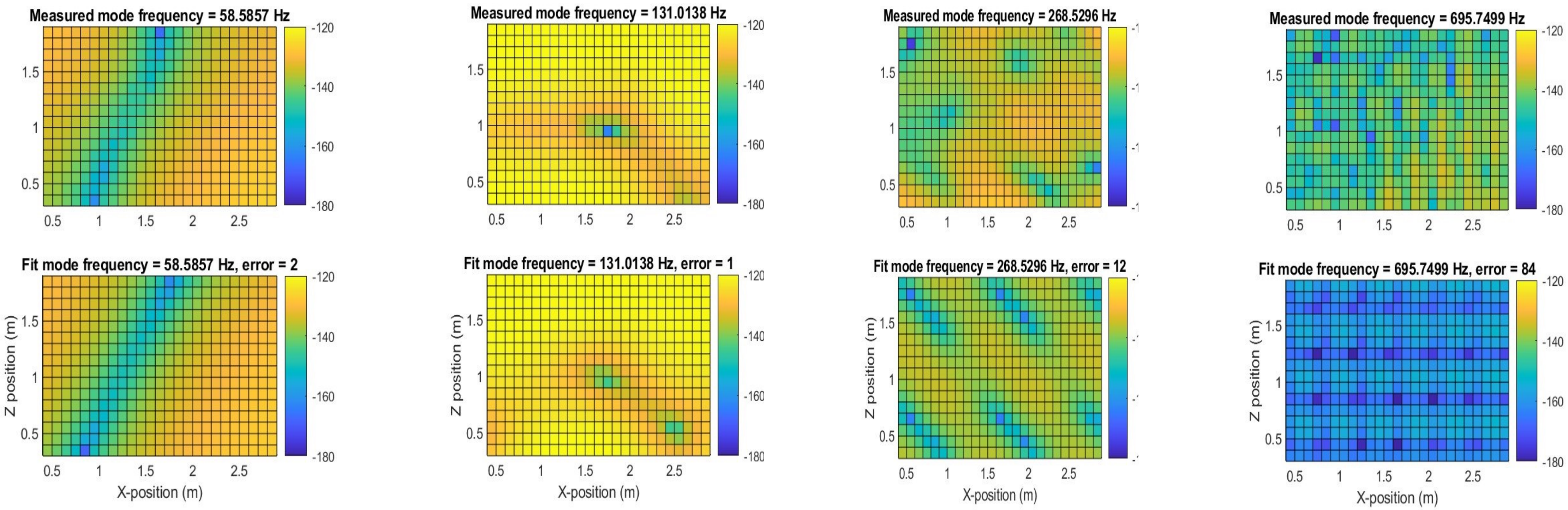
Left and right wall admittance - 0.8

Floor and ceiling admittance - 0.99

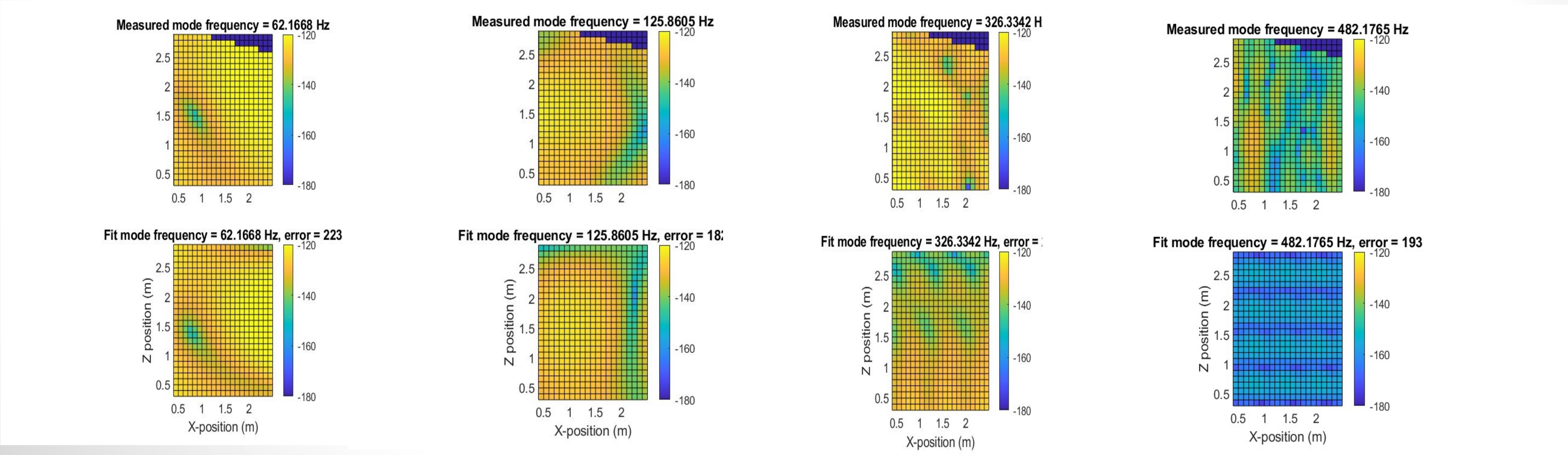


[6] Saarelma, J., Califa, J. and Mehra, R..”Challenges of distributed real-time finite-difference time-domain room acoustic simulation for auralization.” In *2018 AES International Conference on Spatial Reproduction-Aesthetics and Science*. Audio Engineering Society.

Modal fit for shoebox room – 442 mics



Modal fit for room with tilted walls – 594 mics

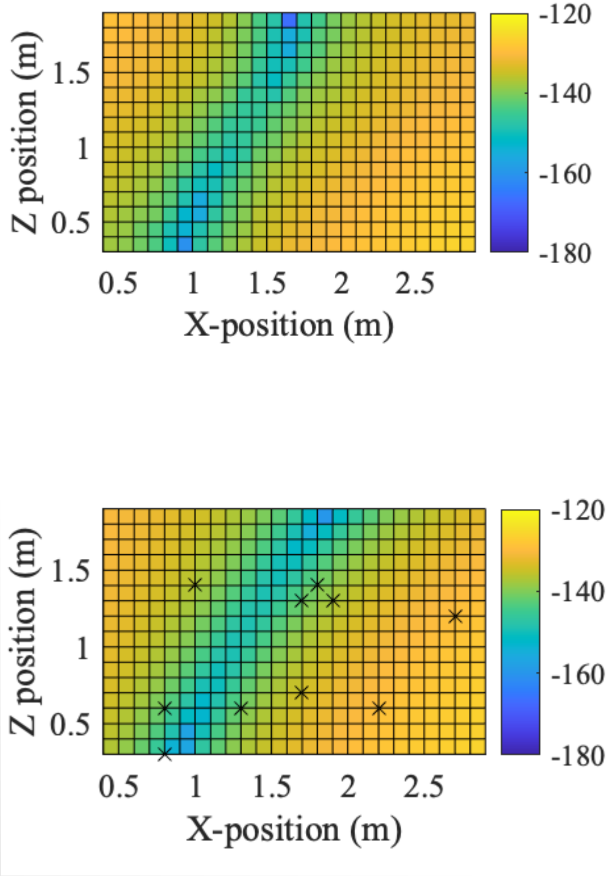


Experiment with Sparse Microphones

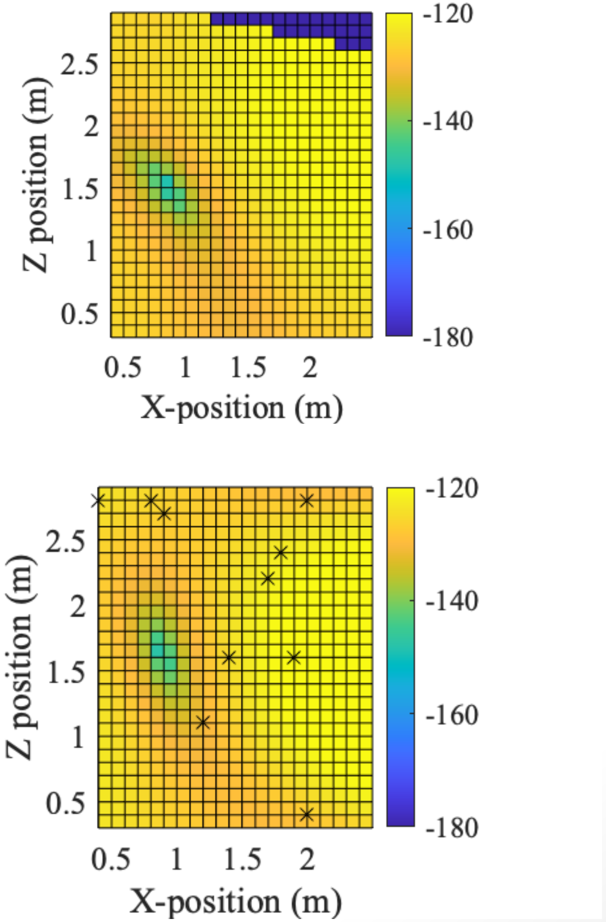
- To replicate sparsity, vary number of microphones in each simulated room from ***5 to 50***.
- Run ***N_{tr} = 100*** trials for each set, placing the microphones in different configurations.
- Calculate and plot
  - ***Mean Structural Similarity Index (MSSIM)*** – metric between [0,1] to measure similarity between actual and fit mode shapes.
  - ***Absolute Mean Spectral Difference Error (AMSDE) -***
    $\frac{1}{N_{tr}}\sum_{n=1}^{N_{tr}} \left|20 \log_{10} \frac{\sum_{l=1}^{L} H_{l,n}(\omega)}{\sum_{l=1}^{L} \widehat{H}_{l,n}(\omega)}\right|$

Results

Modal fits with 10 microphones

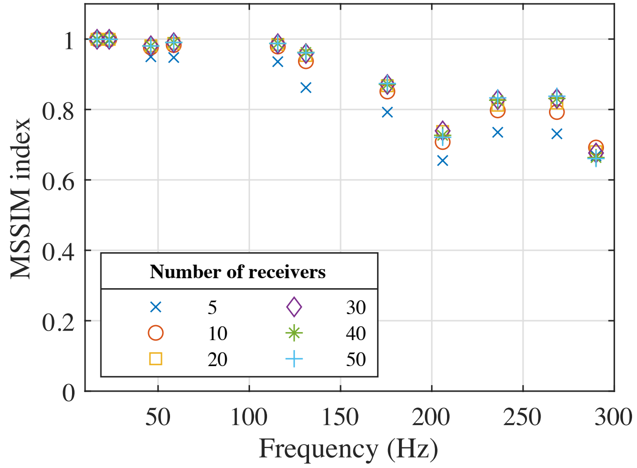


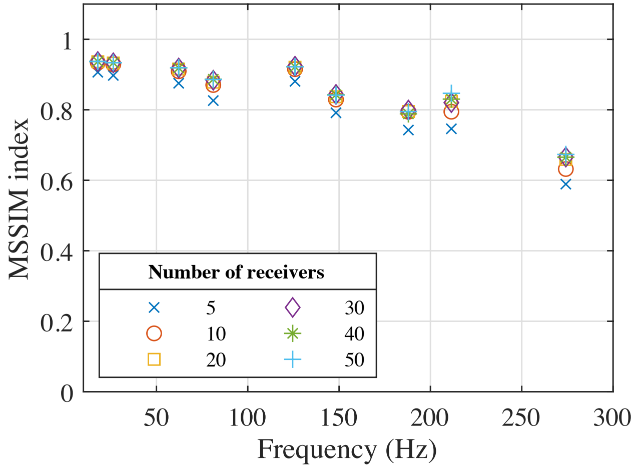
Shoebox, Mode at 58 Hz



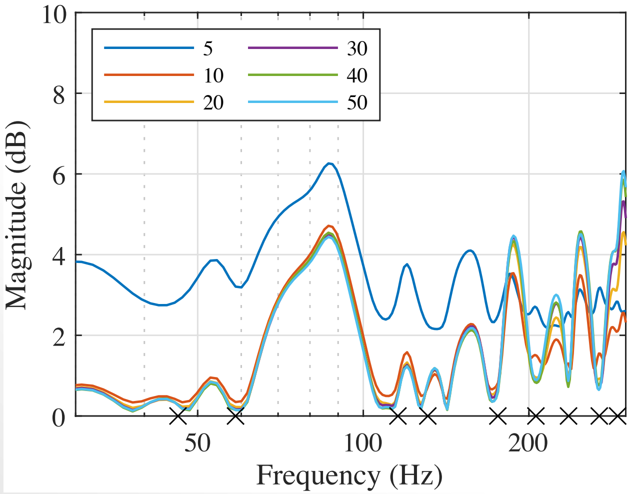
Tilted Walls, Mode at 62 Hz

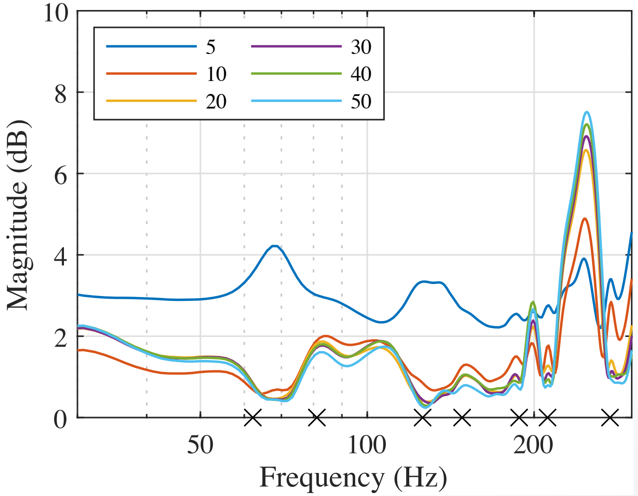
Results





MSSIM





AMSDE

Shoebox

Tilted Walls

Conclusion

- Proposed method is capable of accurate RIR interpolation in the lower frequencies with a very small number of randomly distributed microphones.
- Only a handful of parameters need to be stored for characterizing the low-frequency spatial response of a room - mode frequencies, dampings, two wave numbers and four constants for each mode.
- RIR can be interpolated very efficiently in real-time once these parameters have been estimated offline.

Future Work

- Comparison with State-of-the-art DL method.[7]
- Perceptual evaluation.
- Extend model beyond low frequencies.

[7] Lluís, F., Martínez-Nuevo, P., Bo Møller, M. and Ewan Shepstone, S. "Sound field reconstruction in rooms: Inpainting meets super-resolution." *The Journal of the Acoustical Society of America*, *148*(2), pp.649-659, 2020.

FACEBOOK REALITY LABS

AUDIO