Australian National University

Room Impulse Response Reconstruction Based on Spatio-temporal-spectral Features Learned from a Spherical Microphone Array Measurement

Amy Bastine[1], Thushara D. Abhayapala[1], Jihui (Aimee) Zhang[2,1]

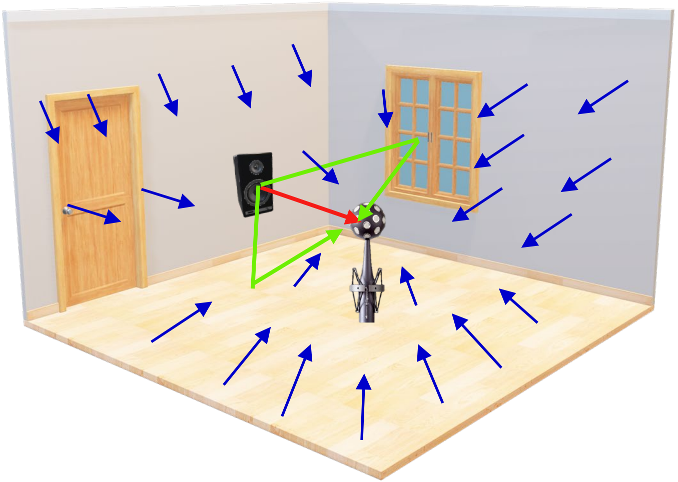
[1]*Audio & Acoustic Signal Processing Group*
College of Engineering, Computing and Cybernetics
The Australian National University
Canberra, Australia

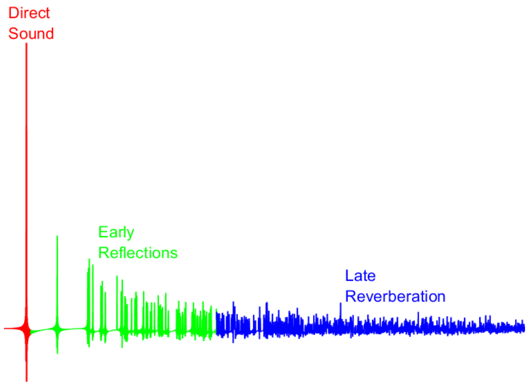
[2]*Signal Processing, Audio and Hearing Group*
Institute of Sound and Vibration Research (ISVR)
University of Southampton
Southampton, United Kingdom

2023 International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
Special Session: Data Driven and Machine Learning based Room Acoustic Modeling

Room Impulse Responses (RIRs)

Fundamental representation of a room acoustic system





- Large-scale RIR measurements required to determine room's response to different source-listener configurations
- **RIR reconstruction methods**
  - Enable estimation of listener experience outside the measurement positions
  - Reduce measurement costs

RIR Reconstruction

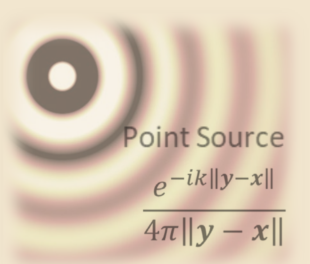
Existing Methods

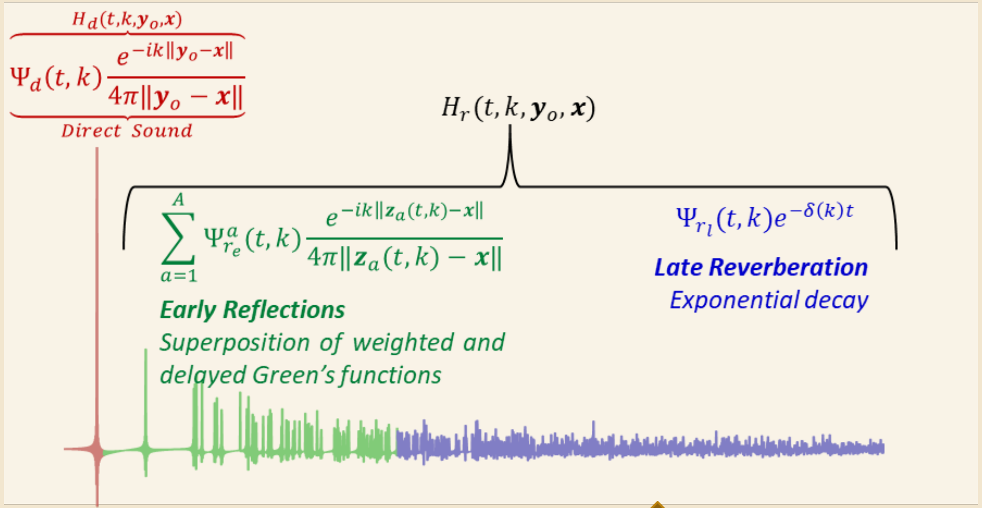
- Model-based methods: Useful for prediction, but may not accurately emulate the intended room response
    - *Wave*-based: Computationally expensive
    - *Geometrical* models like Image-Source Method (ISM), Ray Tracing: Limited to high frequencies
    - *Hybrid* Approaches

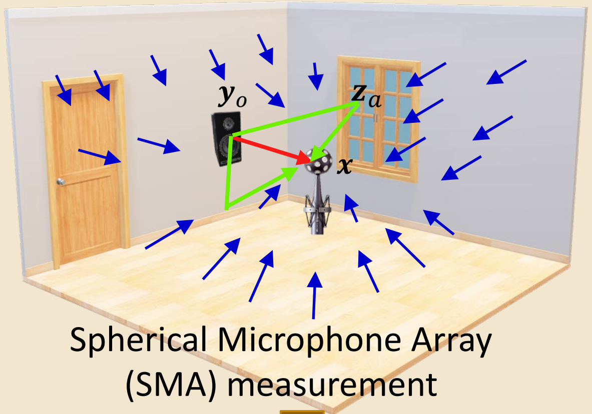
- Data-driven methods based on existing measurements generate more authentic RIRs
    - *Machine learning*: Requires large amounts of training data
    - *Interpolation*: Requires distributed grid of microphone measurements
    - *Extrapolation-based Parametric* methods: Minimal measurement and computational cost
        - Spherical Microphone Arrays (SMAs) & Spherical Harmonics-based processing ⟹ *Higher-order soundfield information ⟹ Improve reconstruction performance*

RIR Reconstruction

Using spatio-temporal-spectral features learned from a SMA measurement







Spherical Microphone Array (SMA) measurement

Features of room reflections:

Reflection source locations: $z_a(t,k)$

Reflection magnitude response: $\Psi_r(t,k)$

Parameter Estimation

Eigenbeam vMF-based Room Acoustic Analyzer

RIR Reconstruction

1. Parameter estimation using *Eigenbeam vMF-based Room Acoustic Analyzer*

RIR Measurements using SMA of radius R and Q microphones $h(\tau, y_o, x_q)$

STFT →

Room Transfer Function $H(t, k, \boldsymbol{y}_o, \boldsymbol{x}_q)$

→

Eigenbeam coefficients

$$\alpha_{nm}(t,k) = \frac{\sum_{q=1}^{Q} \eta_q H(t,k,\boldsymbol{y}_o,\boldsymbol{x}_q)\, Y_{nm}^{*}(\widehat{\boldsymbol{x}_q})}{b_n(kR)}$$

$\{\{n,m\} \in \mathbb{Z} \mid 0 \le n \le N; -n \le m \le n\}$
$N = \lceil kR \rceil$ and $N \le \left\lfloor \sqrt{Q} - 1 \right\rfloor$

↓ *Separate reflections from direct path using* ***Eigenbeam Spatial Correlation*** *Method* (Appendix I)

Estimation of reflection-power coefficients $\Gamma_{vu}(t,k)$

$\{\{v,u\} \in \mathbb{Z} \mid 0 \le v \le V; -v \le u \le v\}$
$V \le \left\lfloor \sqrt{\sqrt{(N+1)^4 - 1} - 1} \right\rfloor$

Reflection Magnitude response
$\sqrt{\Gamma_{00}(t,k)}$

↓

Angular reflection power

$$P_r(t,k,\hat{\boldsymbol{z}}) = \sum_{v=0}^{V} \sum_{u=-v}^{v} \Gamma_{vu}(t,k)\, Y_{vu}(\hat{\boldsymbol{z}})$$

Model fitting using ***Expectation-Maximization (EM) algorithm*** *integrated with Bayesian Minimum Message Length (MML)-based clustering* (Appendix II)

←

Reflection Source parameters:

Dominant Reflection directions: $\boldsymbol{\mu}_a(t,k)$
Relative strength of each dominant reflection direction: $\omega_a(t,k)$
Number of dominant reflection directions: $A(t,k)$

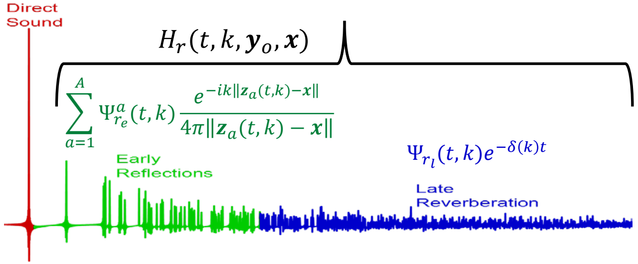
Estimation of directional parameters $\boldsymbol{\mu}_a(t,k)$ and $\omega_a(t,k)$ by fitting 3D von Mises-Fisher (vMF) Mixture Model

$$\mathrm{F}(\mathrm{X}; \boldsymbol{\Pi}_{t,k}) = \sum_{a=1}^{A(t,k)} \omega_a(t,k) \frac{\sqrt{\kappa_a(t,k)}}{(2\pi)^{\left(\frac{3}{2}\right)} I_{\left(\frac{1}{2}\right)}(\kappa_a(t,k))} e^{(\kappa_a(t,k)\boldsymbol{\mu}_a(t,k)^T \mathbf{X})}$$

A. Bastine, T. D. Abhayapala, and J. Zhang, “Time-frequency-dependent directional analysis of room reflections using Eigenbeam processing and von Mises–Fisher clustering,” *J. Acoust. Soc. Amer.*, vol. 151, no. 5, pp. 2916–2930, May 2022.

RIR Reconstruction

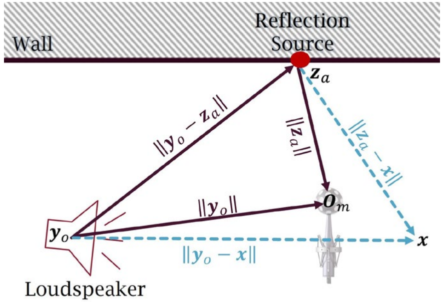
2. Synthesis of Reflection Transfer Function



$$\widetilde{H}_r(t,k,\boldsymbol{y}_o,\boldsymbol{x}) = \begin{cases} \text{Early Reflections } (\textit{superposition of weighted and delayed Green's functions}) & ; \text{for } \{t,k\} \text{ frames where } \boldsymbol{\Pi}_{t,k} \text{ exits} \\ \text{Late Reverberations } (\textit{Exponential Decay}) & ; \text{otherwise} \end{cases}$$

$$= \begin{cases} \sqrt{\Gamma_{00}(t,k)} \sum_{a=1}^{A(t,k)} A(t,k)\omega_a(t,k) \frac{e^{-ik\|\boldsymbol{z}_a(t,k)-\boldsymbol{x}\|}}{4\pi\|\boldsymbol{z}_a(t,k)-\boldsymbol{x}\|} e^{-ikd_t} & ; \text{for } \{t,k\} \text{ frames where } \boldsymbol{\Pi}_{t,k} \text{ exits} \\ \sqrt{\Gamma_{00}(t,k)} \frac{\widetilde{H}_r(t-1,k,\boldsymbol{y}_o,\boldsymbol{x})}{\sqrt{\Gamma_{00}(t-1,k)}} e^{-\delta(k)t_f} e^{-ikd_t} & ; \text{otherwise} \end{cases}$$

- $\sqrt{\Gamma_{00}(t,k)}$: Time-frequency-dependent magnitude response of room reflections
- $A(t,k)\omega_a(t,k)$: Directional amplitude scaling
- $\boldsymbol{z}_a(t,k)$: Location of room surface point in the direction of $\boldsymbol{\mu}_a(t,k)$ *(Approximate room dimensions known)*
- e^{-ikd_t} : Phase-shift to align the response to the corresponding STFT time frame
- d_t : $(t-1)ct_f$ with t_f being the time gap between STFT frames
- $\delta(k)$: Decay rate calculated from $\sqrt{\Gamma_{00}(t,k)}$



RIR Reconstruction

3. Combining Direct and Reflection Components

$$\tilde{H}_d(t,k,\boldsymbol{y}_o,\boldsymbol{x}) = \Psi_d(t,k)\frac{e^{-ik\|\boldsymbol{y}_o-\boldsymbol{x}\|}}{4\pi\|\boldsymbol{y}_o-\boldsymbol{x}\|}$$

ISTFT

$$\tilde{h}_d(\tau,\boldsymbol{y}_o,\boldsymbol{x})$$

Normalize

$$\tilde{H}_r(t,k,\boldsymbol{y}_o,\boldsymbol{x})$$

ISTFT

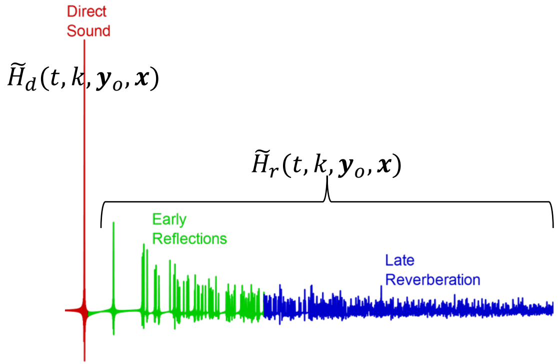
$$\tilde{h}_r(\tau,\boldsymbol{y}_o,\boldsymbol{x})$$

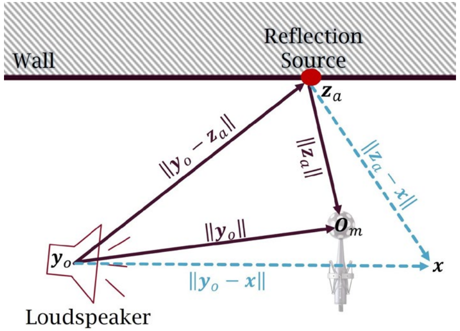
Normalize

$$\tilde{h}(\tau,\boldsymbol{y}_o,\boldsymbol{x}) = \tilde{h}_d(\tau,\boldsymbol{y}_o,\boldsymbol{x}) + \tilde{h}_r\left(\tau - \frac{\overline{\|\boldsymbol{y}_o-\boldsymbol{z}_a(t=1,k)\|}}{c},\boldsymbol{y}_o,\boldsymbol{x}\right)e^{-\frac{\bar{\delta}d_p}{c}}$$

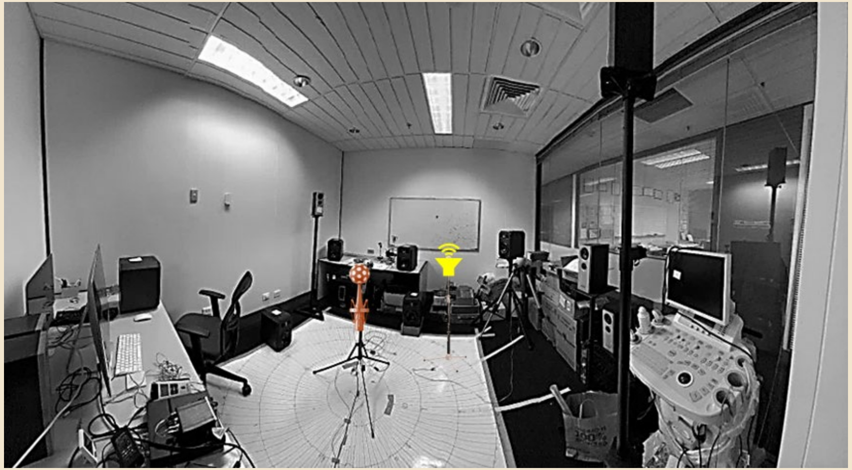
$$d_p = \overline{\|\boldsymbol{y}_o-\boldsymbol{z}_a(t=1,k)\|} + \overline{\|\boldsymbol{z}_a(t=1,k)-\boldsymbol{x}\|}$$

$$\bar{\delta} = \frac{3\ln(10)}{T_{60}}$$





Experimental Analysis

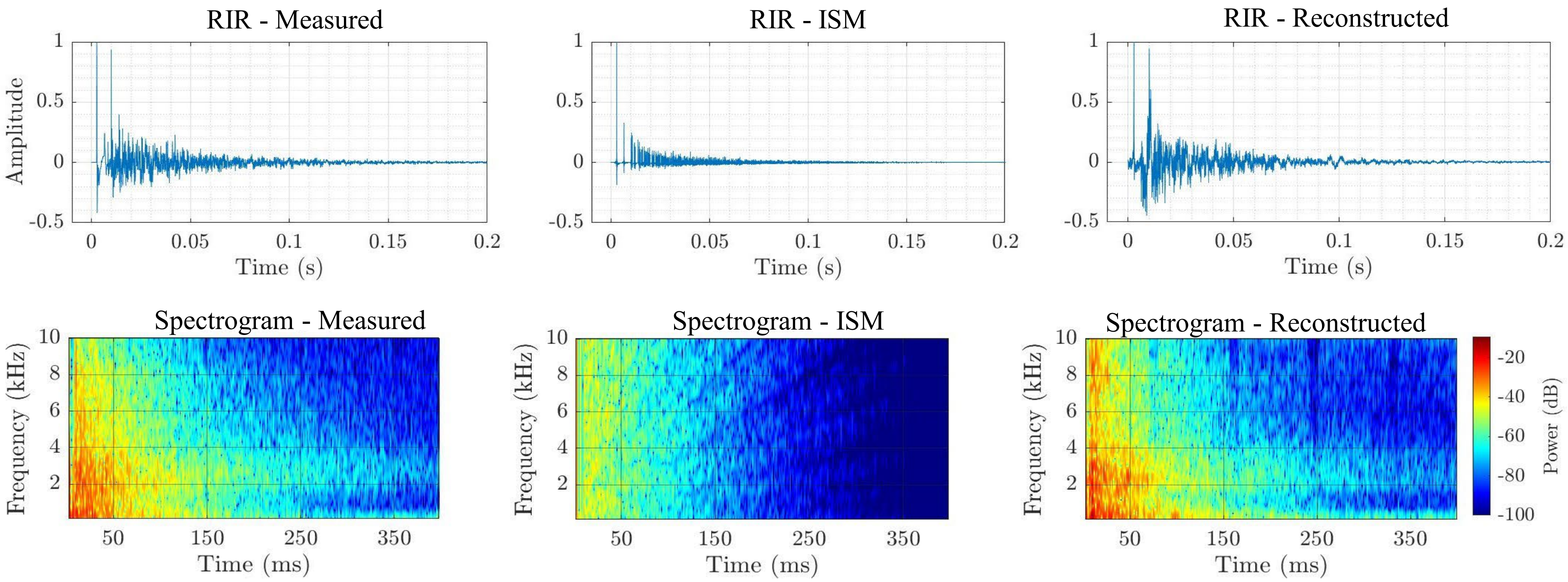


- Room of size $3.54 \times 4.06 \times 2.7$ m and $\mathrm{T}_{60} = 0.329$ s
- Source located at the spherical coordinate $\boldsymbol{y}_o = (1, 90°, 40°)$

- Parameters estimated from RIRs recorded by an EM32 Eigenmike (32-element rigid SMA with radius $\mathrm{R} = 0.042$ m)
  - Maximum number of dominant reflection directions $\left(A(t,k)\right)$ set to 5
  - Total of 7448 reflection sources $\boldsymbol{z}_a(t,k)$ identified from all STFT frames

- Performance compared with measured RIRs and conventional Image Source Method (ISM)
  - ISM-based RIRs generated with maximum image order (Number of image sources in the order of 10^5)

Experimental Analysis

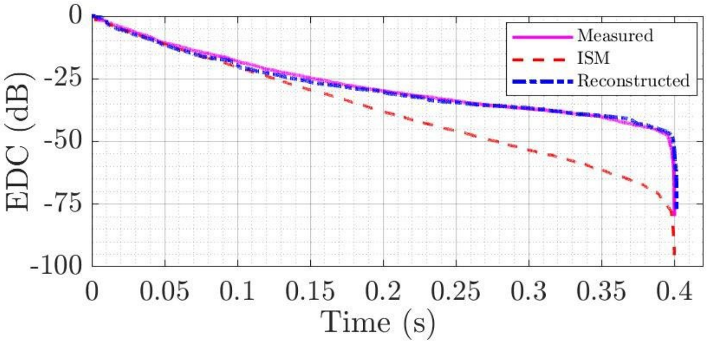


RIRs and corresponding spectrograms for a source located at $\boldsymbol{y}_o = (1{,}90°, 40°)$ and receiver at $x = (0.042, 69°, 0°)$

$\text{NMSE (ISM, Measured)}_{0.4\ \text{s}} = -5\ \text{dB}$

$\text{NMSE (Reconstructed, Measured)}_{0.4\ \text{s}} = -17\ \text{dB}$

Experimental Analysis



Energy Decay Curves (EDC) of RIRs obtained using normalized Schroeder integration method for a source located at $\boldsymbol{y}_o = (1, 90°, 40°)$ and receiver at $x = (0.042, 69°, 0°)$.

Features of early and late reflections preserved in the reconstructed RIRs

Mean and Standard Deviation (STD) of objective room acoustic parameters calculated for the 32 receiver positions of EM32. $e_{I,M}$ and $e_{R,M}$ represent deviation errors of the mean parameter values of ISM-based and Reconstructed RIRs from the measured RIR, respectively.

	Parameters	Measured			ISM			Reconstructed		
		Mean	STD	JND	Mean	STD	$e_{I,M}$	Mean	STD	$e_{R,M}$
Early Decay Time	EDT (s)	0.24	0.007	$5\% = 0.012$	0.14	0.001	0.1	0.23	0.0016	0.01
Clarity	C_{80} (dB)	15.36	0.63	1	17.18	0.38	1.79	16.42	0.22	1.06
Reverberation Time	T_{30} (s)	0.20	0.012	$5\% = 0.01$	0.15	0.0025	0.05	0.20	0.001	0.0
Gravity Time	T_s (ms)	20.69	1.86	10	19.01	1.02	1.89	20.08	0.57	0.61

Reconstructed RIR preserves the perceptual characteristics $\Leftarrow$ Deviations ($e_{R,M}$) are within the Just Noticeable Difference (JND) limits

Current Work

Testing for different rooms

- Room of size $6.5 \times 8.3 \times 2.9$ m and $\mathrm{T}_{60} = 1.12$ s
- $\boldsymbol{y}_o = (1{,}90°, 0°)$
- Reconstructed for a receiver at $\boldsymbol{x} = (0.042, 69°, 0°)$
  - Maximum $A(t, k)$ set to 10
  - Total of 18129 reflection sources $\boldsymbol{z}_a(t, k)$

Parameters	Measured		Reconstructed	
	Value	JND	Value	$e_{R,M}$
EDT (s)	0.36	$5\% = 0.018$	0.35	0.01
C_{80} (dB)	12.57	1	12.23	0.34
T_{30} (s)	0.47	$5\% = 0.0235$	0.47	0.0
T_S (ms)	16.17	10	17.56	1.39

- Room of size $5.75 \times 7.87 \times 2.91$ m and $\mathrm{T}_{60} = 1.2$ s
- $\boldsymbol{y}_o = (1.8, 88°, 56°)$
- Reconstructed for a receiver at $\boldsymbol{x} = (1.29, 87°, 0°)$
  - Maximum $A(t, k)$ set to 5
  - Total of 9465 reflection sources $\boldsymbol{z}_a(t, k)$

Parameters	Measured		Reconstructed	
	Value	JND	Value	$e_{R,M}$
EDT (s)	0.35	$5\% = 0.0175$	0.38	0.03
C_{80} (dB)	13.55	1	13.36	0.19
T_{30} (s)	0.33	$5\% = 0.0165$	0.32	0.01
T_S (ms)	16.34	10	17.53	1.19

Conclusion

- *Reconstructed RIRs successfully preserved the temporal and spectral behaviors of the measured RIRs*
  - Achieved with single SMA measurement and few parameters per time-frequency frame
    - Dominant directions of reflections and their relative weights
    - Reflection magnitude response
  - Limitation: Approximate dimensions of the room should be known

- *Future Works*:
  - Incorporating angular spread (κ_a) of reflections from room surfaces
  - Develop the method to facilitate listener translations
  - Perceptual evaluation

Thank you for your attention !

Questions?

Australian National University

Appendix I

Parameter estimation using *Eigenbeam vMF-based Room Acoustic Analyzer*

Eigenbeam coefficients

$$\alpha_{nm}(t,k) = \frac{\sum_{q=1}^{Q} \eta_q H(t,k,\boldsymbol{y}_o,\boldsymbol{x}_q)\, Y_{nm}^*(\widehat{\boldsymbol{x}_q})}{b_n(kR)}$$

→ Estimation of reflection-power coefficients $\Gamma_{vu}(t,k)$

Separate reflections from direct path component

$\alpha_{nm}(t,k)$ → Eigenbeam Correlation Matrix $\boldsymbol{\Lambda}(t,k) = \mathrm{E}\{\boldsymbol{\alpha}(t,k)\ \boldsymbol{\alpha}^H(t,k)\}$ → $\widehat{\boldsymbol{\Omega}}(t,k) = \boldsymbol{B}^\dagger(k)\boldsymbol{\Lambda}(t,k)$ → Reflection−power coefficients: $\Gamma_{vu}(t,k)$; **Reflection magnitude response**: $\sqrt{\Gamma_{00}(t,k)}$

$$\boldsymbol{B}(k) = \begin{bmatrix} \delta_{0000} & d_{000000} & \cdots & d_{0000VV} \\ \delta_{001-1} & d_{001-100} & \cdots & d_{001-1VV} \\ \vdots & \vdots & \vdots & \vdots \\ \delta_{00NN} & d_{00NN00} & \cdots & d_{00NNVV} \\ \delta_{1-100} & d_{1-10000} & \cdots & d_{1-100VV} \\ \vdots & \vdots & \vdots & \vdots \\ \delta_{NNNN} & d_{NNNN00} & \cdots & d_{NNNNVV} \end{bmatrix} \quad \boldsymbol{\Lambda}(t,k) = \begin{bmatrix} \Lambda_{0000} \\ \Lambda_{001-1} \\ \vdots \\ \Lambda_{00NN} \\ \Lambda_{1-100} \\ \vdots \\ \Lambda_{NNNN} \end{bmatrix} \quad \boldsymbol{\Omega}(t,k) = \begin{bmatrix} P_D \\ \Gamma_{00} \\ \Gamma_{1-1} \\ \vdots \\ \Gamma_{V-V} \\ \vdots \\ \Gamma_{VV} \end{bmatrix}$$

P_D → Direct Path Power

Γ_{00} → **Reflection power response** $\propto$ $|H_r(t,k,\boldsymbol{y}_o,\boldsymbol{x})|$ (*Direction-independent*)

$\Gamma_{vu}(k,t)$: Reflection Power Coefficients

$\delta_{nmn'm'}$: $16\pi^2 i^{(n-n')} Y_{nm}^*(\widehat{\boldsymbol{y}_o}) Y_{n'm'}(\widehat{\boldsymbol{y}_o})$

$d_{nmn'm'vu}$: $16\pi^2 i^{(n-n')}(-1)^m \sqrt{\frac{(2v+1)(2n+1)(2n'+1)}{4\pi}} \begin{pmatrix} v & n & n' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v & n & n' \\ u & -m & m' \end{pmatrix}$

$\Lambda_{nmn'm'}$: $\mathrm{E}\{\alpha_{nm}(t,k)\alpha_{n'm'}^*(t,k)\}$

Parameter estimation using *Eigenbeam vMF-based Room Acoustic Analyzer*

Angular reflection power
$P_r(t,k,\hat{\mathbf{z}}) = \sum_{v=0}^{V}\sum_{u=-v}^{v}\Gamma_{vu}(t,k)\,Y_{vu}(\hat{\mathbf{z}})$

→

Estimation of directional parameters $\boldsymbol{\mu}_a(t,k)$ and $\omega_a(t,k)$ by fitting 3D von Mises-Fisher (vMF) Mixture Model
$\mathrm{F}(\mathrm{X};\boldsymbol{\Pi}_{t,k}) = \sum_{a=1}^{A(t,k)}\omega_a(t,k)\frac{\sqrt{\kappa_a(t,k)}}{(2\pi)^{\left(\frac{3}{2}\right)} I_{\left(\frac{1}{2}\right)}(\kappa_a(t,k))} e^{(\kappa_a(t,k)\boldsymbol{\mu}_a(t,k)^T\mathbf{X})}$

Angular reflection power
$P_r(t,k,\hat{\mathbf{z}}) = \sum_{v=0}^{V}\sum_{u=-v}^{v}\Gamma_{vu}(t,k)\,Y_{vu}(\hat{\mathbf{z}})$

→

Transform $P_r(t,k,\hat{\mathbf{z}})$ into a distribution on a unit sphere
Statistical frequency of unit vectors in $\hat{\mathbf{z}}$ direction $\propto |P_r(t,k,\hat{\mathbf{z}})|$

→

Estimation of **directional parameters $\boldsymbol{\Pi}_{k,t}$** by fitting 3D vMF Mixture Model $\mathrm{F}(\mathfrak{X};\boldsymbol{\Pi}_{t,k})$ using
Expectation- Maximization (EM) algorithm integrated with Bayesian Minimum Message Length (MML)-based clustering

$$\mathrm{F}(\mathfrak{X};\boldsymbol{\Pi}_{t,k}) = \sum_{a=1}^{A(t,k)}\omega_a(t,k)\frac{\sqrt{\kappa_a(t,k)}}{(2\pi)^{\left(\frac{3}{2}\right)} I_{\left(\frac{1}{2}\right)}(\kappa_a(t,k))}\exp(\kappa_a(t,k)\boldsymbol{\mu}_a(t,k)^T\mathfrak{X})$$

- $\mathfrak{X}$: Set of uniformly sampled points on the surface of a unit sphere
- $\boldsymbol{\mu}_a(t,k)$: Mean Direction Vector ⇒ *Dominant Reflection Directions* of $\{t,k\}^{th}$ bin
- $\omega_a(t,k)$: Convex mixing coefficients ⇒ *Relative strength of each $\boldsymbol{\mu}_a$ direction*
- $A(t,k)$: Number of vMF components ⇒ *Number of dominant reflection directions*
- $\kappa_a(t,k)$: Non-negative Concentration Parameter ⇒ *Dispersion of reflected power from $\boldsymbol{\mu}_a$*
- $\boldsymbol{\Pi}_{t,k}$: $\{\boldsymbol{\mu}_1,\cdots,\boldsymbol{\mu}_A;\omega_1,\cdots,\omega_A;\ \kappa_1,\cdots,\kappa_A\}$ ⇒ *Detectable only for $\{t,k\}$ bins with anisotropic (early) reflections*