Aalto University
School of Electrical
Engineering

Similarity Metrics for Late Reverberation

Gloria Dal Santo, Karolina Prawda, Sebastian J. Schlecht, Vesa Välimäki

58th Asilomar Conference 2024, Pacific Grove, CA, USA, 27-30 October 2024

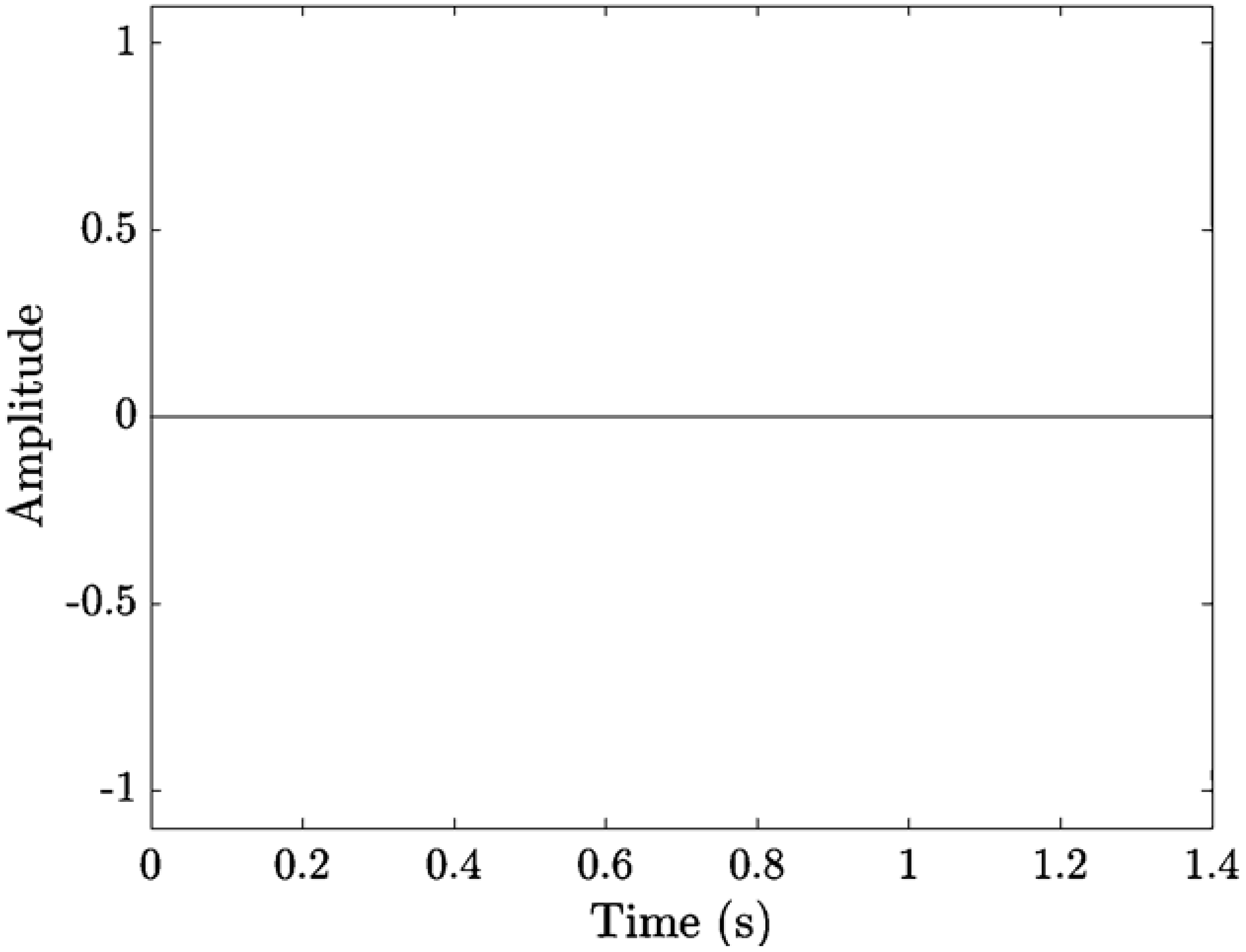
gloria.dalsanto@aalto.fi

Room Impulse Response

RIR: Time-domain representation of acoustic propagation from one source to one receiver inside a room

Artificial Reverb
synthesis of room acoustics

- virtual acoustics
- speech processing
- music production



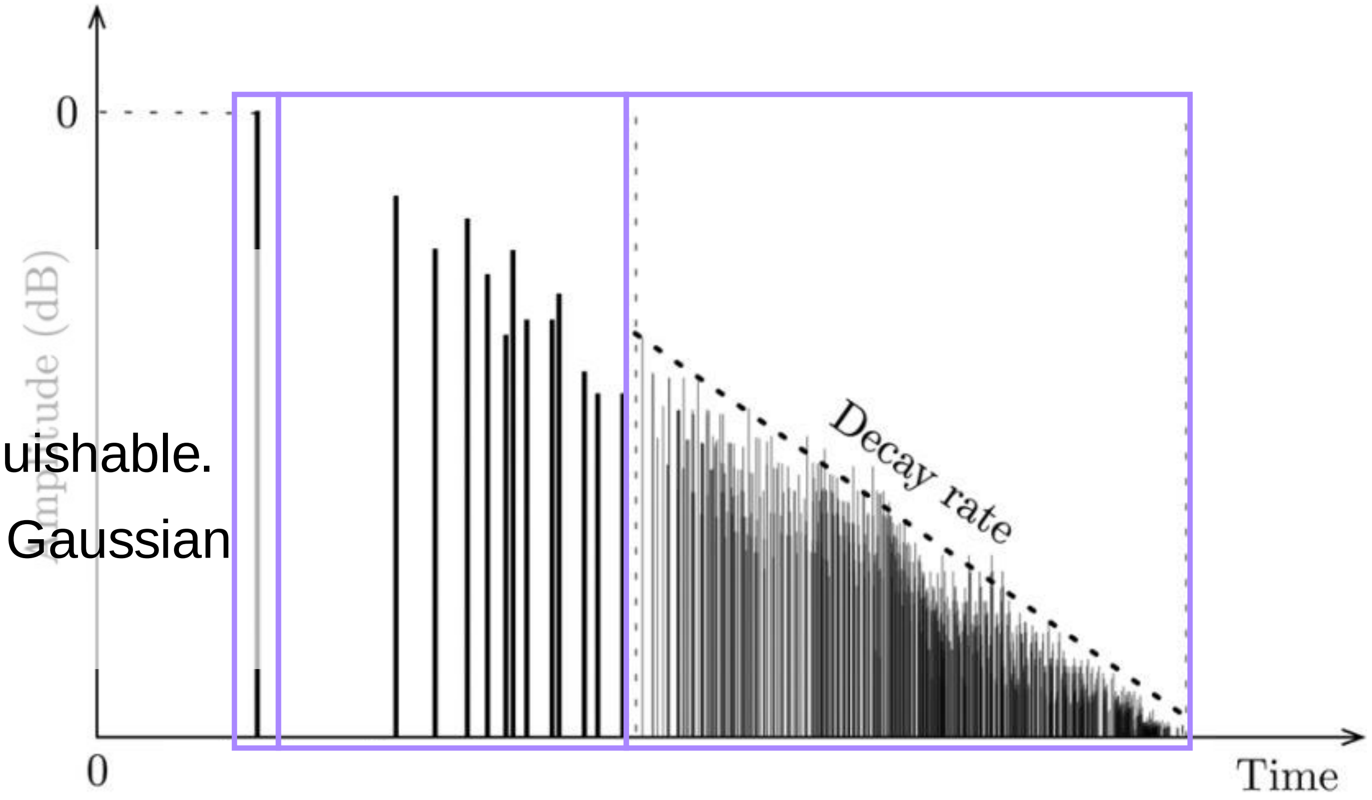
Room Impulse Response

Direct sound

Early reflections

Late reverberation

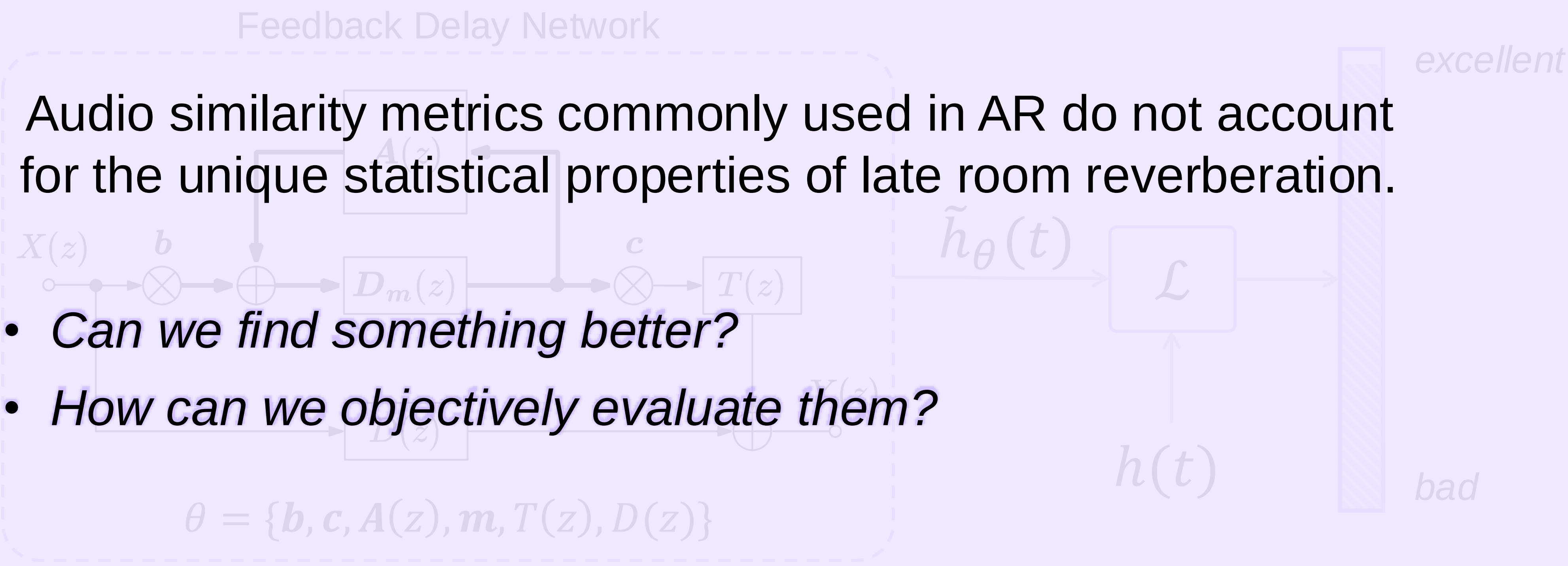
- Sufficiently diffuse sound field.
- Individual reflections are indistinguishable.
- Energy ~ exponentially decaying Gaussian noise.



Välimäki, Vesa, et al. "Fifty years of artificial reverberation." *IEEE Transactions on Audio, Speech, and Language Processing* 20.5 2012

Artificial Reverberation

Automatic tuning of parametric artificial reverberation algorithms to match a target room impulse response $h(t)$ based on a cost function $\mathcal{L}(\tilde{h}_\theta, h)$ 'LineWidth',2.5);



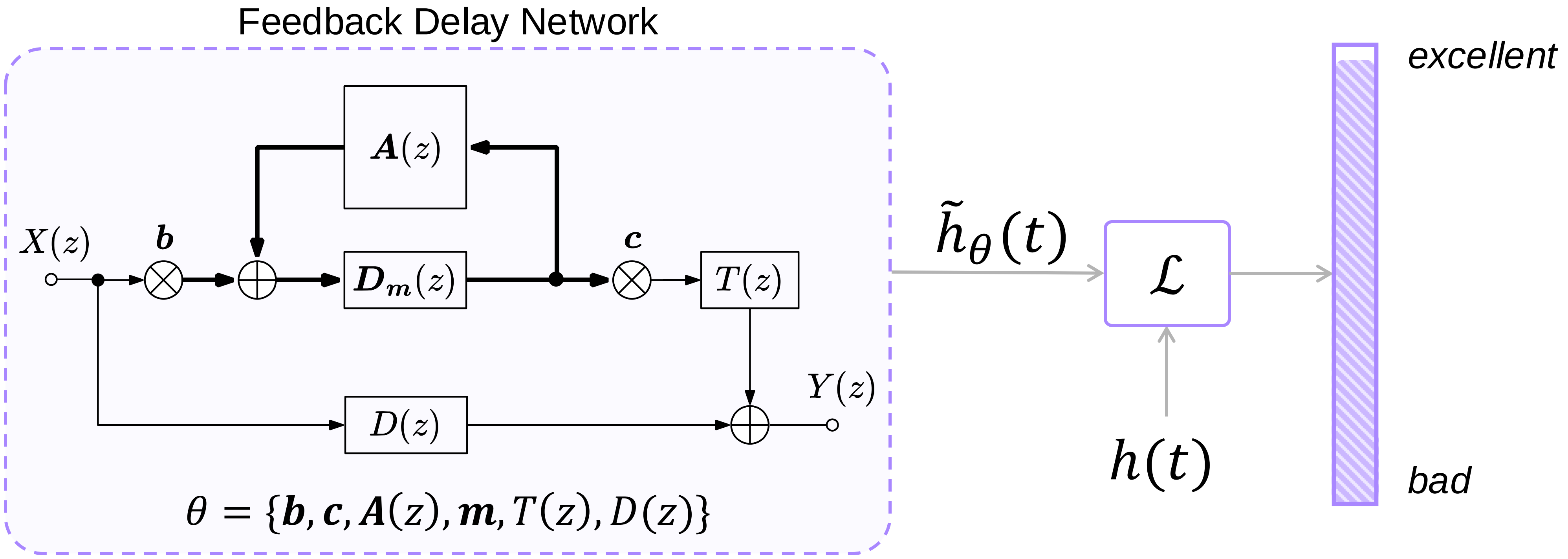
Audio similarity metrics commonly used in AR do not account for the unique statistical properties of late room reverberation.

- *Can we find something better?*
- *How can we objectively evaluate them?*

Dal Santo, Gloria, et al. "RIR2FDN: An improved room impulse response analysis and synthesis." in Proc. DAFx 2024.

Artificial Reverberation

Automatic tuning of parametric artificial reverberation algorithms to match a target room impulse response $h(t)$ based on a cost function $\mathcal{L}(\tilde{h}_\theta, h)$ 'LineWidth',2.5);



Dal Santo, Gloria, et al. "RIR2FDN: An improved room impulse response analysis and synthesis." in Proc. DAFx 2024.

Proposed Metrics Averaged Power Convergence

Distance on the time-frequency representation:

$$\mathcal{L}_{\mathrm{PC}} = \left\| \frac{|H(t,f)|^2 * W - |\hat{H}(t,f)|^2 * W}{(|H(t,f)|^2 * W)(|\hat{H}(t,f)|^2 * W)} \right\|_{\mathrm{F}}$$

$\boldsymbol{H}(t,f)$ Short-time Fourier transform
$\boldsymbol{W}$ 2D Hann window
$||\cdot||_F$ Frobenius norm

- Difference in local time-frequency averaged power.
- Mitigates the effect of short-term fluctuations.
- Expected to converge for similar reverberation conditions

Proposed Metrics Energy Decay Convergence

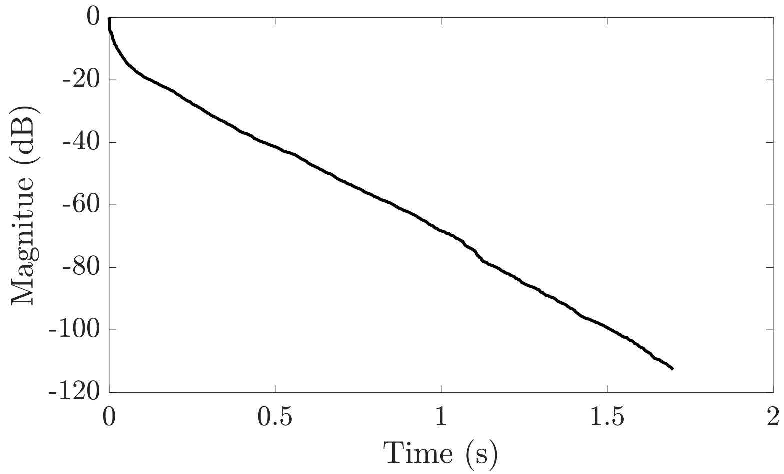
Convergence of the energy level over time and frequency:

$$\mathcal{L}_{\mathrm{EDC}} = \frac{1}{|\mathcal{C}|} \sum_{f_\mathrm{c} \in \mathcal{C}} \frac{\sum_{t=0}^{L} (\varepsilon_{\mathrm{dB}}(t; f_\mathrm{c}) - \hat{\varepsilon}_{\mathrm{dB}}(t; f_\mathrm{c}))^2}{\sum_{t=0}^{L} \varepsilon_{\mathrm{dB}}^2(t; f_\mathrm{c})}$$

$\varepsilon_{\mathrm{dB}}(t; f_\mathrm{c})$ energy decay curve at frequency band with center frequency f_c

- Convergence of the level of energy over time.
- $\varepsilon_{\mathrm{dB}}(t; f_\mathrm{c})$ computed from Schroeder backward integration

$$\varepsilon(t; f_\mathrm{c}) = \sum_{\tau=t}^{L} h_{f_\mathrm{c}}^2(\tau)$$



Proposed Metrics Baselines

Multi Scale Spectral Loss

$$\mathcal{L}_{\mathrm{MSS}}(h,\hat{h}) = \frac{1}{M}\sum_{m=1}^{M}(\mathcal{L}_{\mathrm{SC}}(h,\hat{h}) + \mathcal{L}_{\mathrm{SM}}(h,\hat{h}))$$

Error-to-signal ratio

$$\mathcal{L}_{\mathrm{ESR}}(h,\hat{h}) = \frac{\sum_{t_{\mathrm{mix}}}^{L}|h(t)-\hat{h}(t)|^2}{\sum_{t_{\mathrm{mix}}}^{L}|h(t)|^2}.$$

spectral matching

$$\mathcal{L}_{\mathrm{SM}}(h,\hat{h}) = \frac{1}{N}\|\log(|H(t,f)|) - \log(|\hat{H}(t,f)|)\|_1$$

spectral convergence

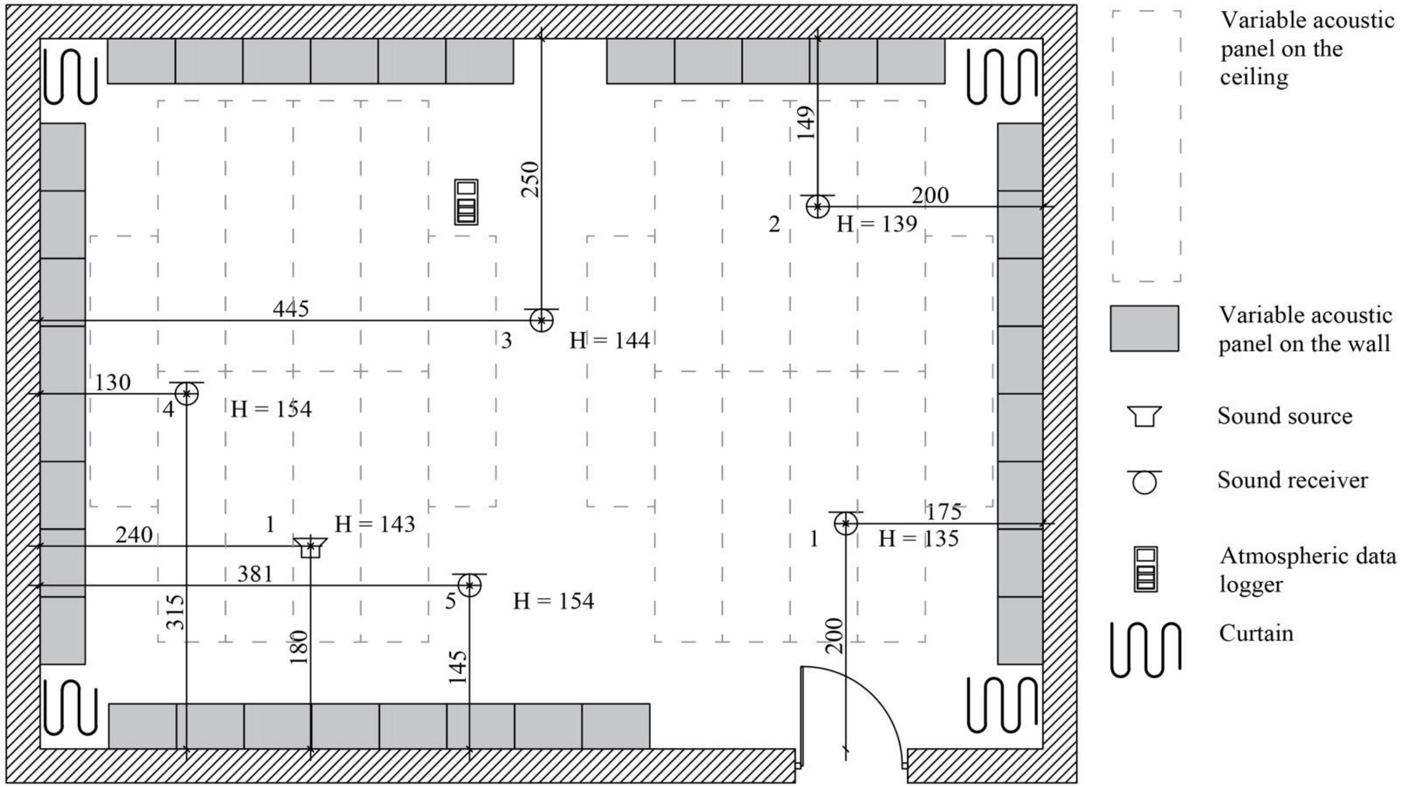
$$\mathcal{L}_{\mathrm{SC}}(h,\hat{h}) = \frac{\||H(t,f)| - |\hat{H}(t,f)|\|_F}{\||H(t,f)|\|_F}$$

R. Yamamoto, et al. "Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram," in Proc. IEEE ICASSP, 2020
A. Wright, et al. "Real-time black-box modelling with recurrent neural networks". In Proc. DAFx, 2019.

Variable Acoustics Room Dataset

- 55 variable acoustics panels
- 5342 recorded panel configurations
- 5 microphone positions
- $t > t_{\text{mix}}$

smooth transitions between
different reverberation conditions

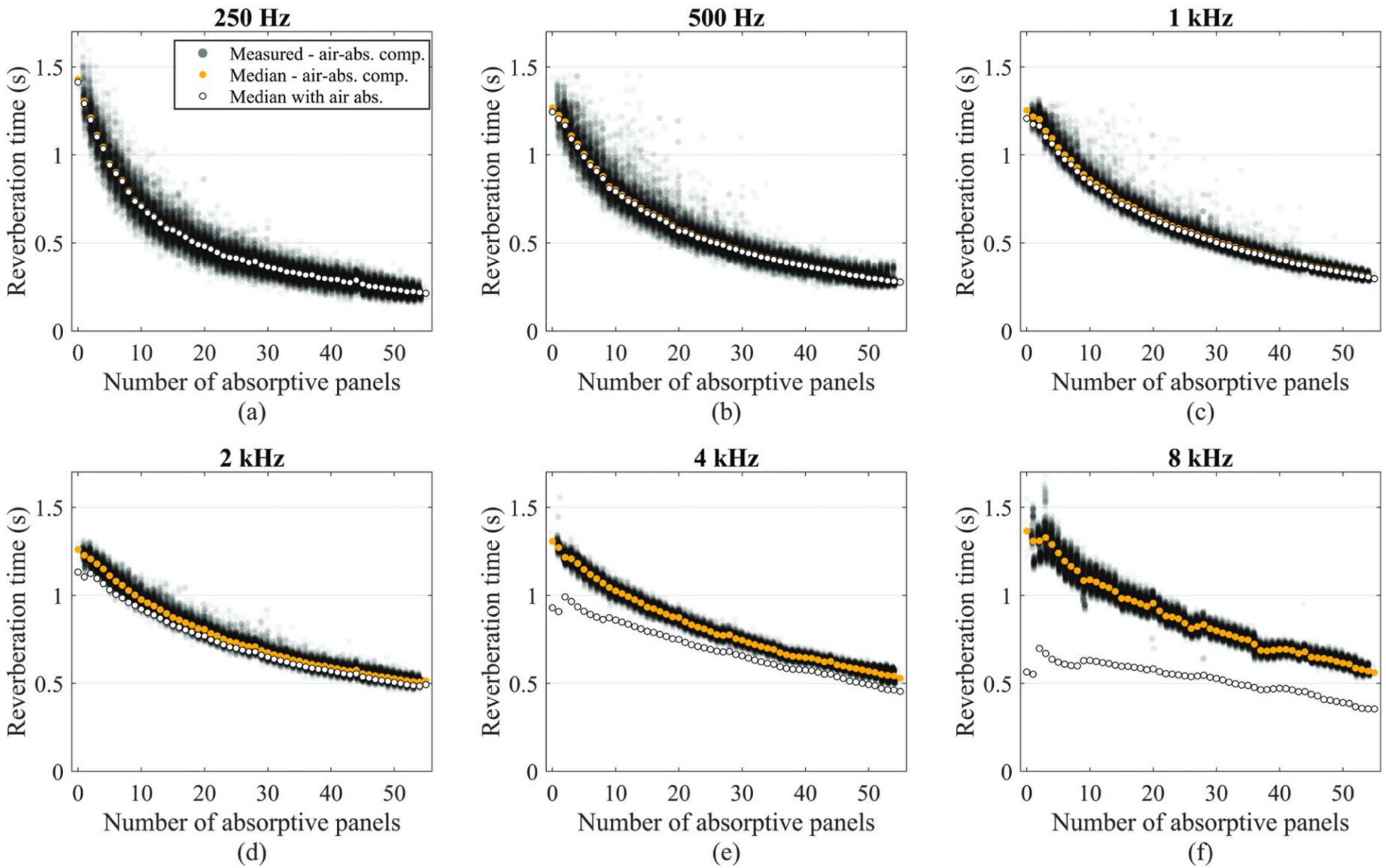


K. Prawda, S. J. Schlecht, and V. Välimäki, “Calibrating the Sabine and Eyring formulas,” J. Acoustical Society of America, 2022.

Variable Acoustics Room Dataset

- 55 variable acoustics panels
- 5342 recorded panel configurations
- 5 microphone positions
- $t > t_{\text{mix}}$

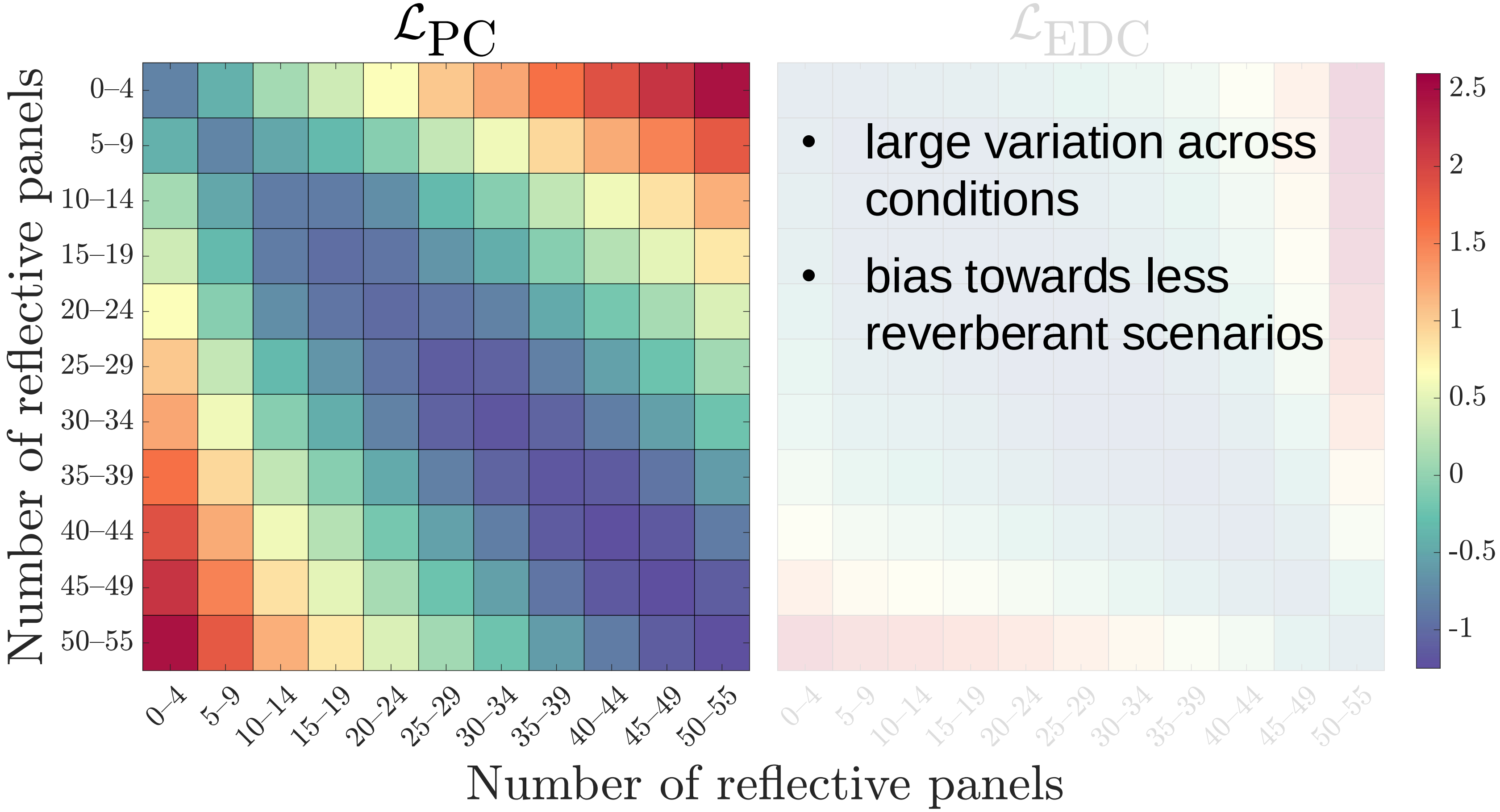
smooth transitions between different reverberation conditions



K. Prawda, S. J. Schlecht, and V. Välimäki, “Calibrating the Sabine and Eyring formulas,” J. Acoustical Society of America, 2022.

Evaluation Reverberation Conditions

Median of loss values after normalization to $\mu = 0$ and $\sigma = 1$

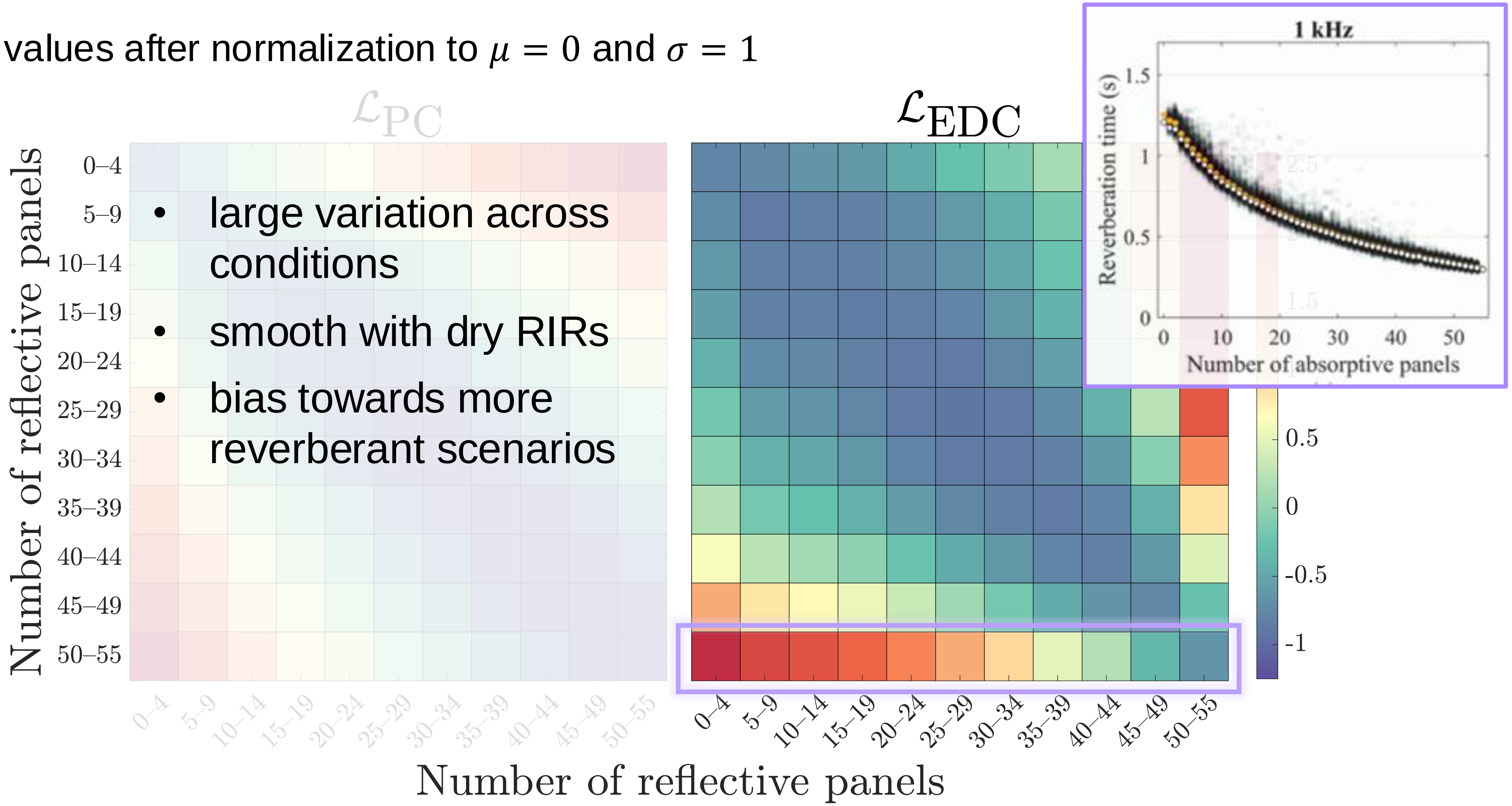


- large variation across conditions
- bias towards less reverberant scenarios

K. Prawda, S. J. Schlecht, and V. Välimäki, “Calibrating the Sabine and Eyring formulas,” J. Acoustical Society of America, 2022.

Evaluation Reverberation Conditions

Median of loss values after normalization to $\mu = 0$ and $\sigma = 1$

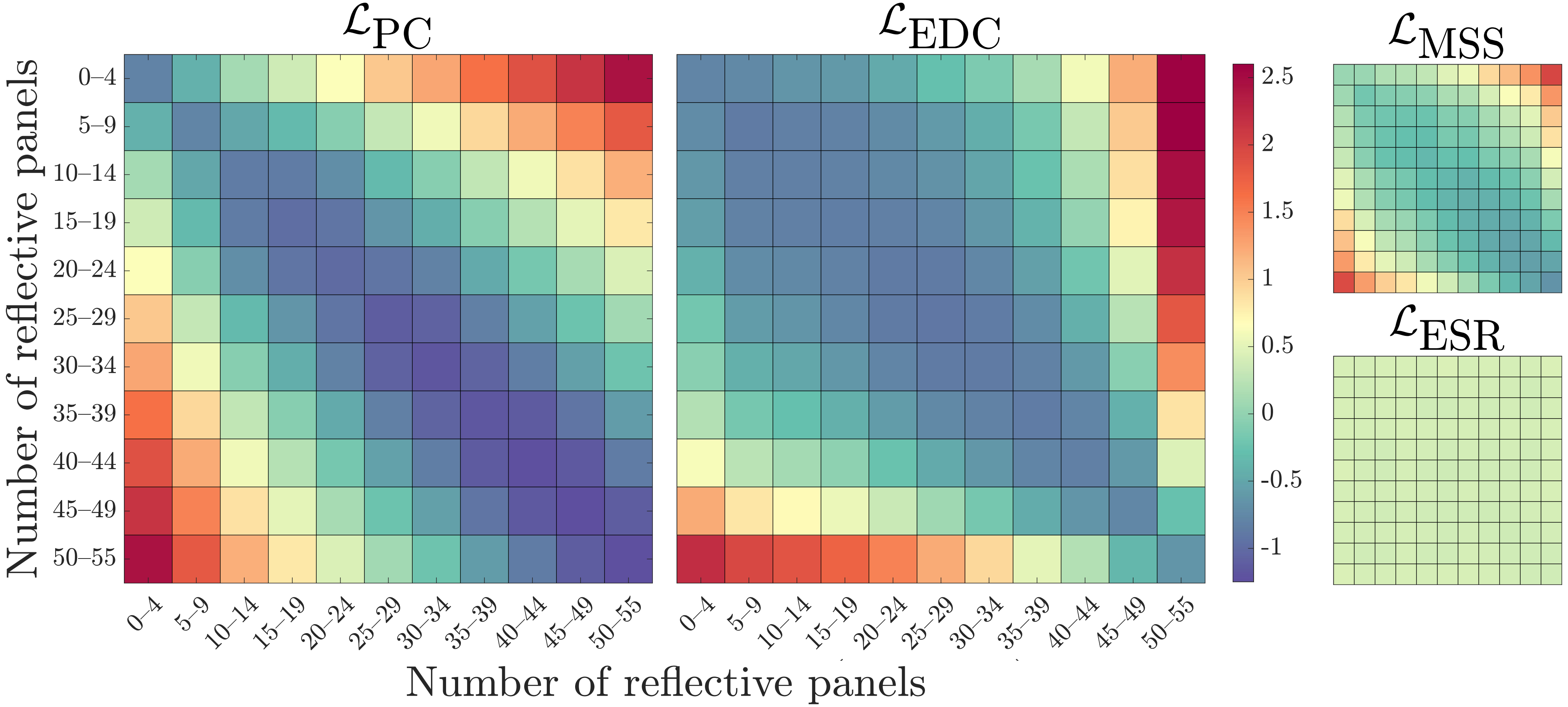


- large variation across conditions
- smooth with dry RIRs
- bias towards more reverberant scenarios

K. Prawda, S. J. Schlecht, and V. Välimäki, “Calibrating the Sabine and Eyring formulas,” J. Acoustical Society of America, 2022.

Evaluation Reverberation Conditions

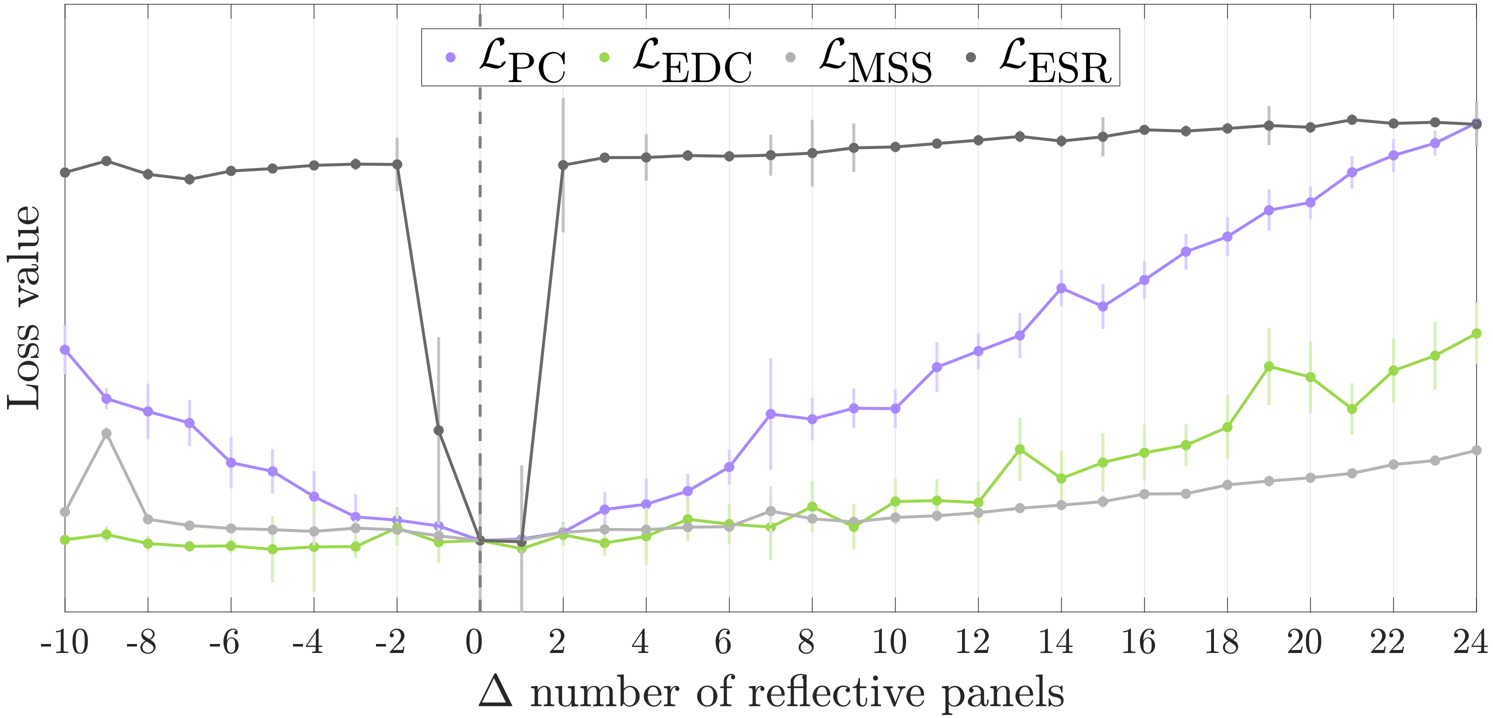
Median of loss values after normalization to $\mu = 0$ and $\sigma = 1$



K. Prawda, S. J. Schlecht, and V. Välimäki, "Calibrating the Sabine and Eyring formulas," J. Acoustical Society of America, 2022.

Evaluation Reverberation Conditions

Distance between the target RIR and 50 randomly chosen RIRs for every number of reflective panels and same microphone position

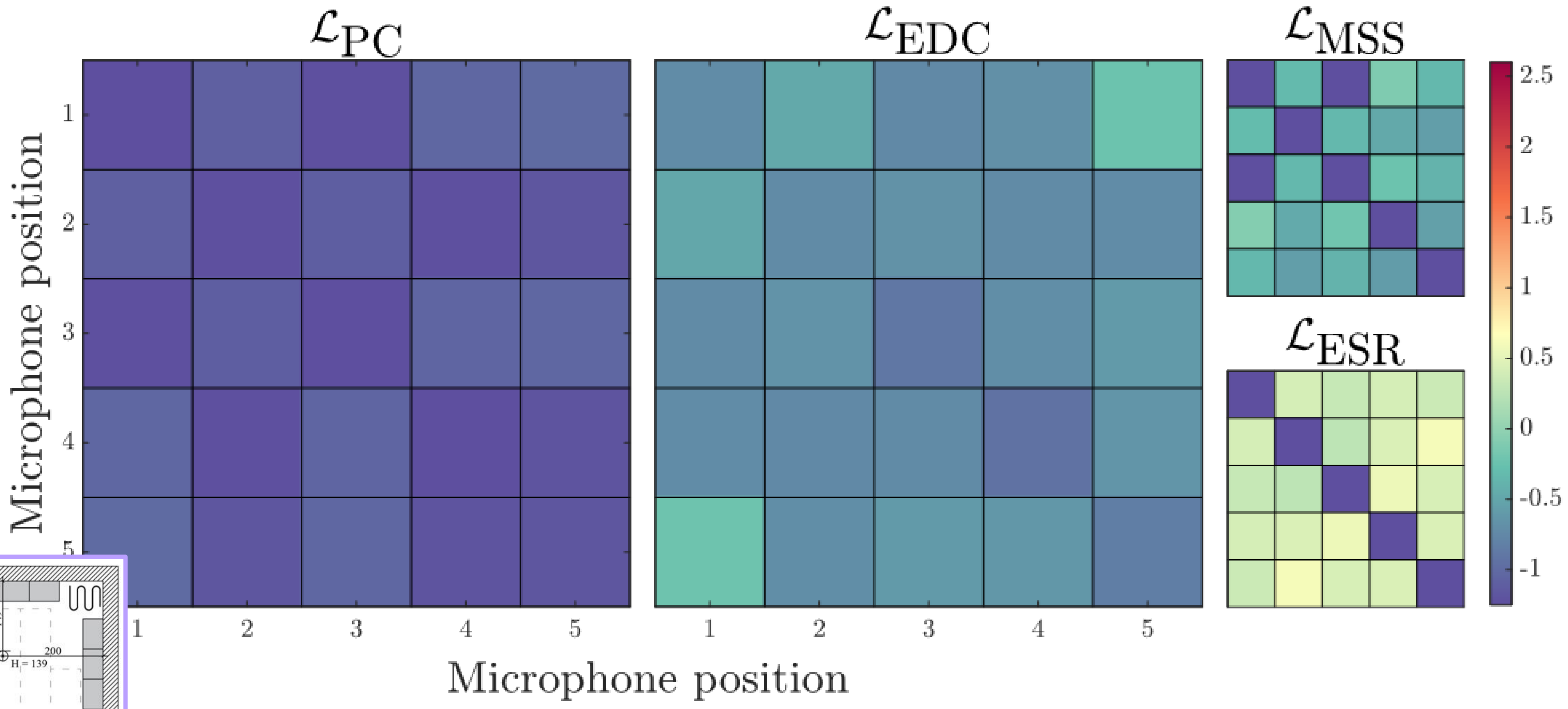


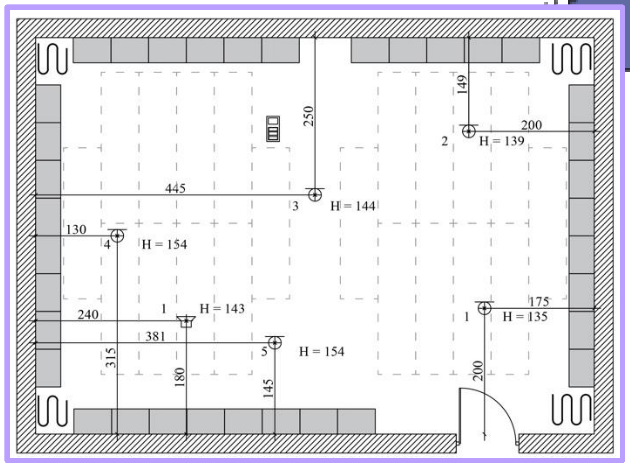
$\mathcal{L}_{\mathrm{PC}}$ shows the most optimal decay towards the minimum, making it ideal ML applications.

K. Prawda, S. J. Schlecht, and V. Välimäki, “Calibrating the Sabine and Eyring formulas,” J. Acoustical Society of America, 2022.

Evaluation Microphone Position

Subset 35-49 closed panels





- $\mathcal{L}_{\mathrm{PC}}$ and $\mathcal{L}_{\mathrm{EDC}}$ better generalize over microphone position

Conclusions and Future Work

- Objective evaluation of RIR similarity metrics via variable acoustics room measurements.
- $\mathcal{L}_{\mathrm{PC}}$ and $\mathcal{L}_{\mathrm{EDC}}$ outperform baselines in capturing acoustic features and are more robust to receiver positions.
- $\mathcal{L}_{\mathrm{PC}}$ shows the most optimal decay towards the minimum.
- $\mathcal{L}_{\mathrm{EDC}}$ is biased towards more reverberant conditions, reflecting the exponential increase in reverberation time with more reflective surfaces.

Future work

- Validate the results using subjective tests.
- Validation in machine learning applications, e.g. tuning of artificial reverb.

Aalto University
School of Electrical
Engineering

Thank you for your attention
Questions?

Code available online