Similarity Metrics for Late Reverberation

Aalto University
School of Electrical Engineering

Gloria Dal Santo[1], Karolina Prawda[1], Sebastian J. Schlecht[2], and Vesa Välimäki[1]
[1]Aalto University, Acoustics Lab, Dept. of Information and Communications Engineering
[2]FAU, Multimedia Communications and Signal Processing

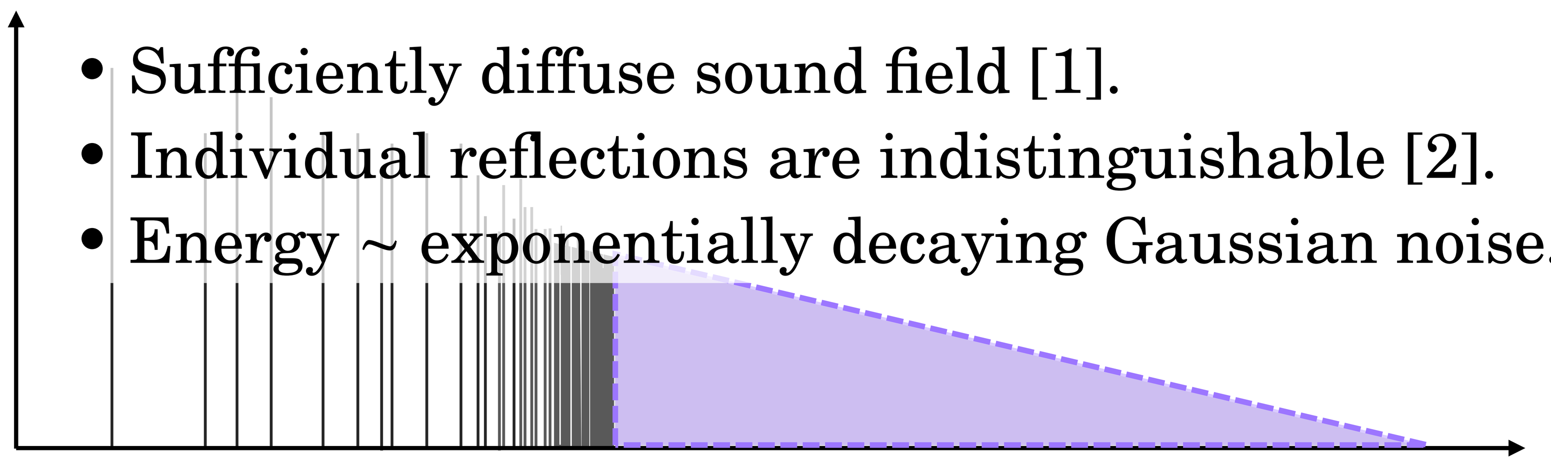
58th Asilomar Conference 2024 Pacific Grove, CA, USA, 27-30 October 2024

github.com/gdalsanto/similarity-metrics-for-rirs

Introduction

Accurate tuning of parametric artificial reverberators relies heavily on the choice of cost function, yet common audio similarity metrics do not account for the unique statistical properties of late room reverberation. We introduce two novel similarity metrics specifically designed for this purpose, which outperform existing metrics on a dataset of measured room impulse responses (RIRs).

Late Reverberation



- Sufficiently diffuse sound field [1].
- Individual reflections are indistinguishable [2].
- Energy ~ exponentially decaying Gaussian noise.

Proposed Similarity Metrics

Averaged power convergence

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

Difference in local time-frequency averaged power.
Mitigates the effect of short-term fluctuations.

Energy decay convergence

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

Convergence of the energy level over time and frequency.

Baselines

The proposed losses are compared to the multi-scale spectral loss (MSS) [3] and the error-to-signal ratio (ESR).

Dataset of measured RIRs

- Variable acoustics room with 55 reflective panels [4].
- 11 subsets of 25 RIRs, with 5 RIRs for each of the 5 receiver positions, based on absorption configuration.

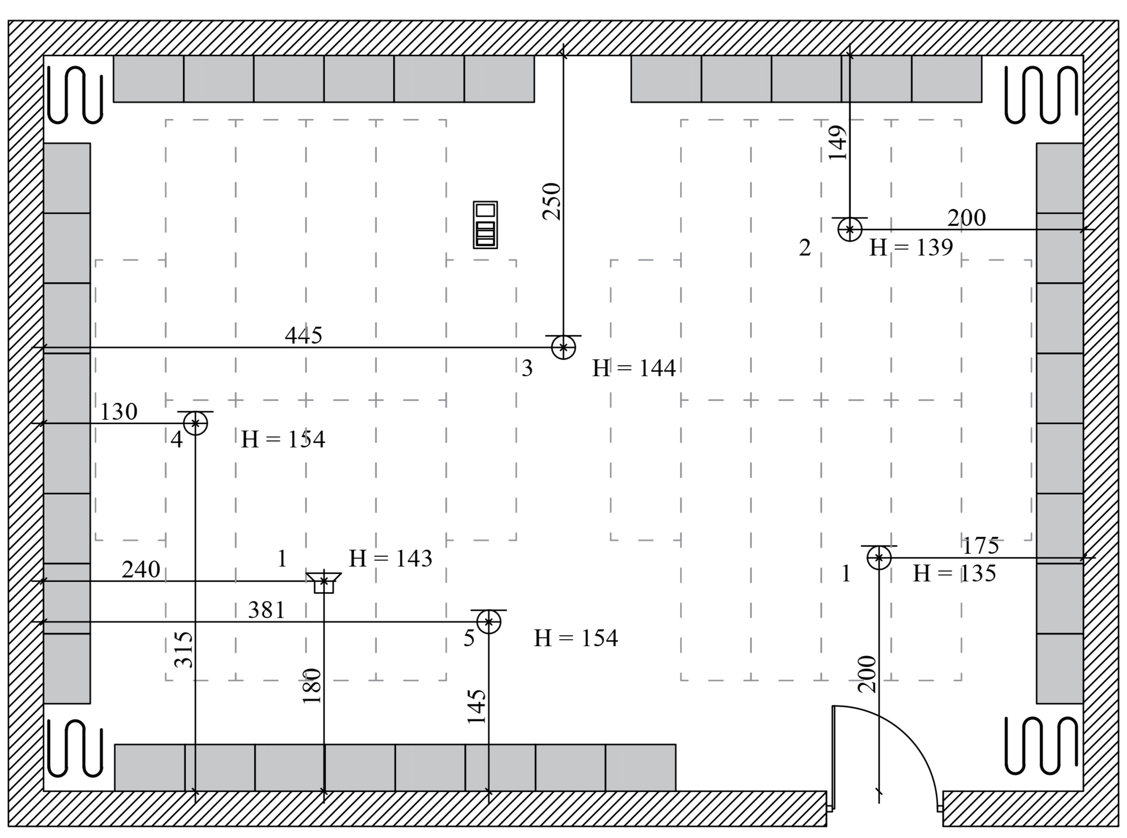


Figure 1. *Layout of the variable acoustics room showing the positions of the panels, sound source, and receivers [4].*

Objective Evaluation

Variations to reverberation conditions

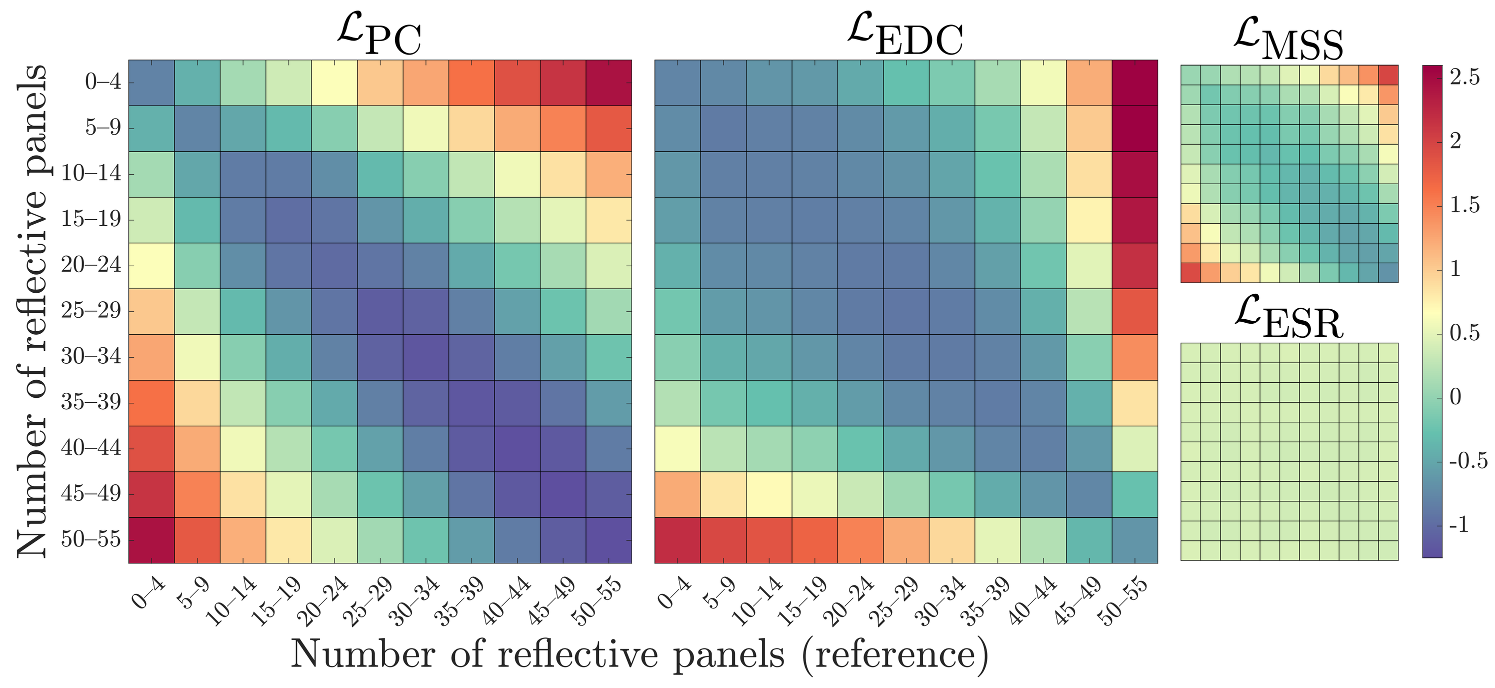


Figure 2. *Median values of the standardized similarity metric distribution for each pair of reflective panel conditions.*

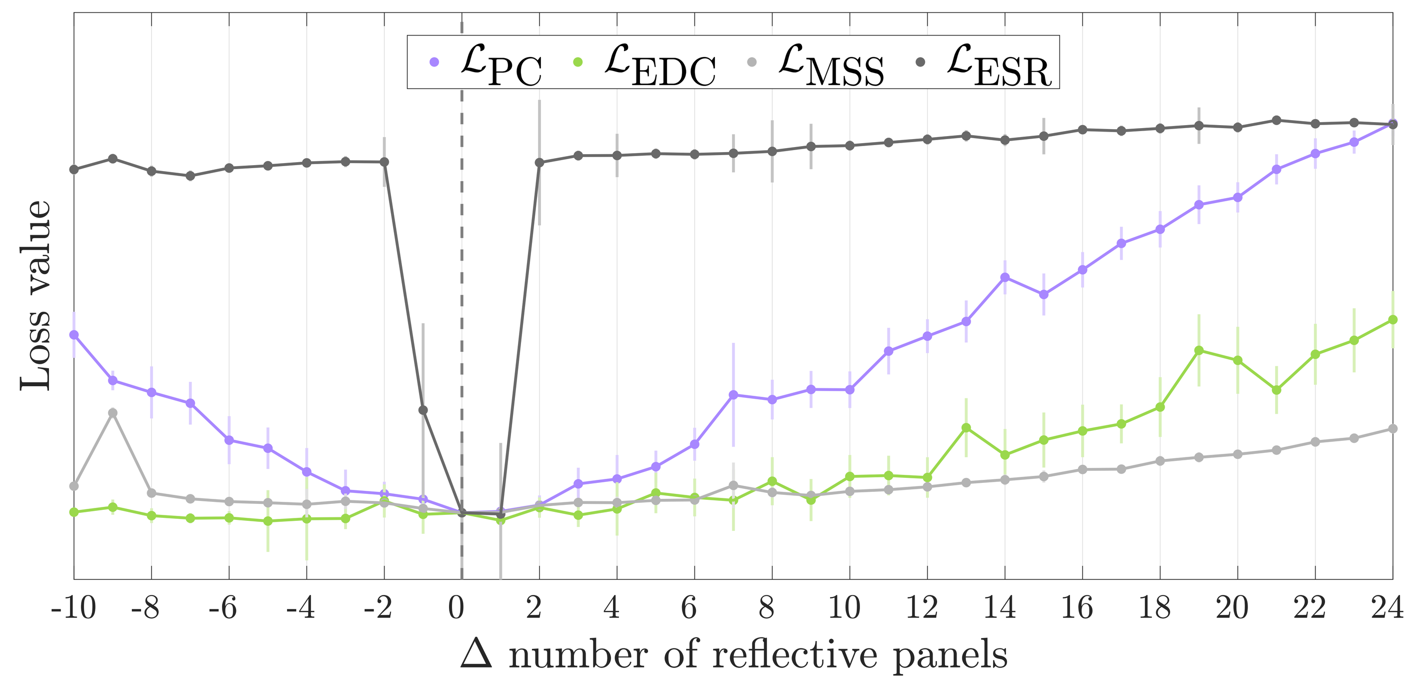


Figure 3. *Evolution of metrics on gradual differences Δ in the number of panels set to reflective position. Medians and standard deviations are marked with dots and vertical lines, respectively. The dashed line indicates the reference RIR's configuration.*

Generalization over receiver location

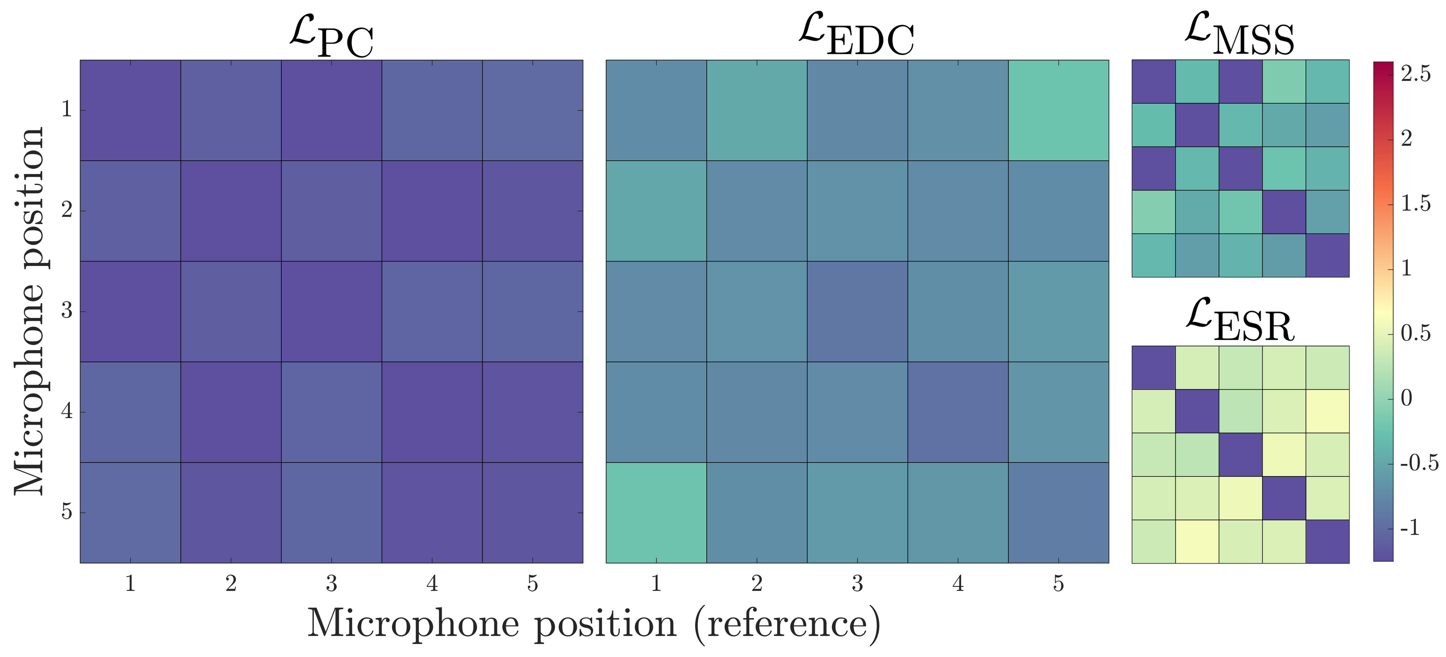


Figure 4. *Median values of the standardized similarity metric distribution for each pair of microphone positions.*

Conclusions

- $\mathcal{L}_{\mathrm{PC}}$ 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, making it ideal for Machine Learning applications.
- $\mathcal{L}_{\mathrm{EDC}}$ is biased towards more reverberant conditions, reflecting the exponential increase in reverberation time with more reflective surfaces.

References

[1] H. Kuttruff, *Room acoustics*. CRC Press, 2016.

[2] J. A. Moorer, "About this reverberation business," *Computer Music J.*, 1979.

[3] R. Yamamoto, E. Song, and J.-M. Kim, "Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram," in *Proc. IEEE ICASSP*, 2020.

[4] K. Prawda, S. J. Schlecht, and V. Välimäki, "Calibrating the Sabine and Eyring formulas," *J. Acoust. Soc. Am.*, 2022.