Aalto-yliopisto
Aalto-universitetet
Aalto University

ELEC-E5631 - Acoustics Seminar

Research towards efficient VR audio engines: Heuristic and data-driven approaches

Georg Götz

In today’s lecture you will learn

How do audio engines work? How to use data-driven methods to simulate and render room acoustics in a plausible way?

Game audio

VR/AR technology

Who of you…

- … owns a VR headset?
- … plays computer games regularly?

Please think about:

- **What makes audio rendering convincing for you** in these application?
- **What is important for you** considering the audio experience**?**

A!

Audio engines need to work in highly dynamic environments

Sound sources and receiver can move and rotate

- ➢ **Source-receiver configuration** comprises both positions and orientations

Environments can change, e.g. player can open/close doors or destroy parts of the environment

A!

The rendering pipeline in audio engines involves multiple modules

	Source	Medium	Receiver
Module	Source →	Medium →	Receiver
Tasks	• Model source directivity • Dynamic rendering: source can move and rotate • Synthesize speech and sounds	• Model acoustic spaces: ○ Sound propagation ○ Sound absorption ○ Sound transmission • Dynamic rendering: environment changes	• Model spatial hearing: ○ HRTFs ○ Simpler models • Dynamic rendering: receiver can move and rotate
Required information	• Directivity pattern • Source-receiver configuration	• Room geometry (3D mesh) • Acoustic material properties • Source-receiver configuration	• HRTFs • Source-receiver configuration

Savioja, L., Huopaniemi, J., Lokki, T. & Väänänen, R. Creating Interactive Virtual Acoustic Environments. *Journal of the Audio Engineering Society* **47**, 675–705 (1999).

The rendering pipeline in audio engines involves multiple modules

Module	Source →	Medium →	Receiver

Tasks	• Model source directivity
	• Dynamic rendering: source can move and rotate
	• Synthesize speech and sounds
		• Model acoustic spaces:
		  ○ Sound propagation
		  ○ Sound absorption
		  ○ Sound transmission
		• Dynamic rendering: environment changes
			• Model spatial hearing:
			  ○ HRTFs
			  ○ Simpler models
			• Dynamic rendering: receiver can move and rotate

Required information	• Directivity pattern • Source-receiver configuration	• Room geometry (3D mesh) • Acoustic material properties • Source-receiver configuration	• HRTFs • Source-receiver configuration

Savioja, L., Huopaniemi, J., Lokki, T. & Väänänen, R. Creating Interactive Virtual Acoustic Environments. *Journal of the Audio Engineering Society* **47**, 675–705 (1999).

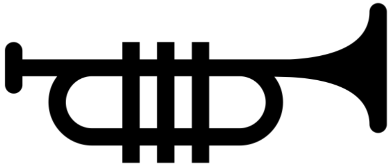
Aalto-yliopisto
Aalto-universitetet
Aalto University

Sound-source modelling

Sound sources have different directivity patterns

Speech or singing voice

Monson, B. B., Hunter, E. J. & Story, B. H. Horizontal directivity of low- and high-frequency energy in speech and singing. *J. Acoust. Soc. Am.* **132**, 433–441 (2012).



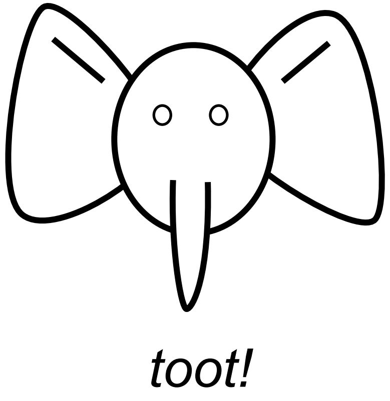
Musical instruments

Pätynen, J. & Lokki, T. Directivities of Symphony Orchestra Instruments. *Acta Acust United Acust* **96**, 138–167 (2010).

Ackermann, D., Brinkmann, F. & Weinzierl, S. A Database with Directivities of Musical Instruments. *J. Audio Eng. Soc.* **72**, 170–179 (2024).

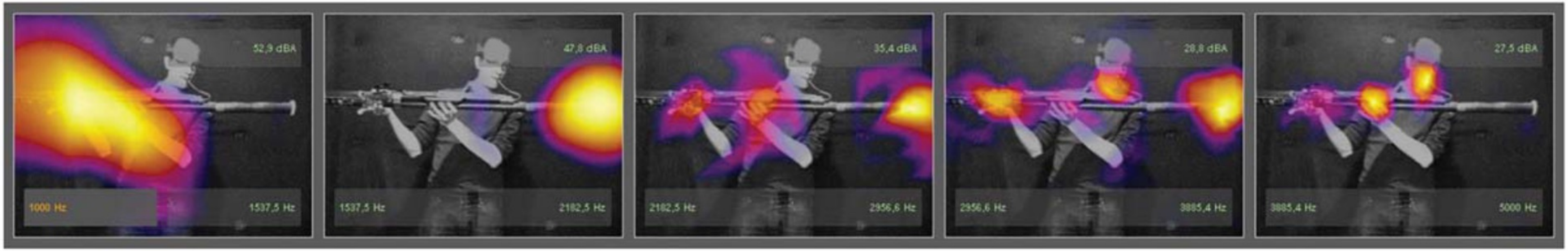
Impulsive sources

Campbell, S., Wall, A., Taylor, C., Mobley, F., & Rasband, R. Large-scale anechoic characterization of small caliber firearm impulse noise. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (pp. 111-120) (2022)



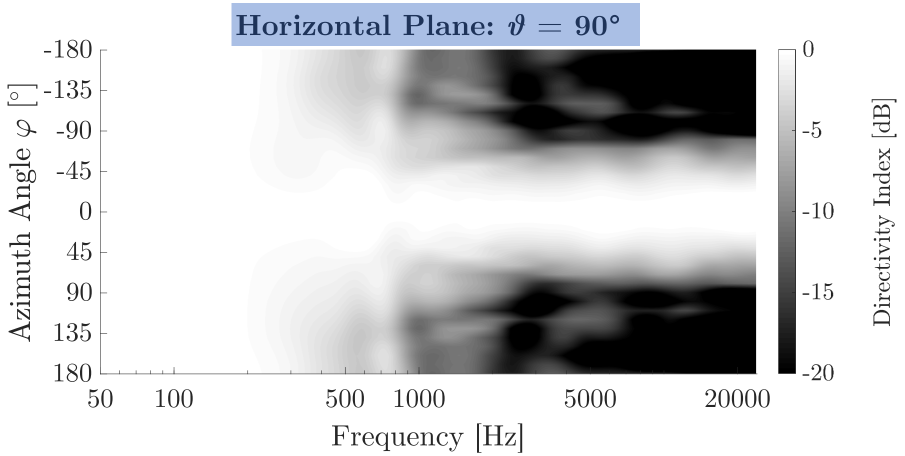
Other frequency-dependent directivity patterns

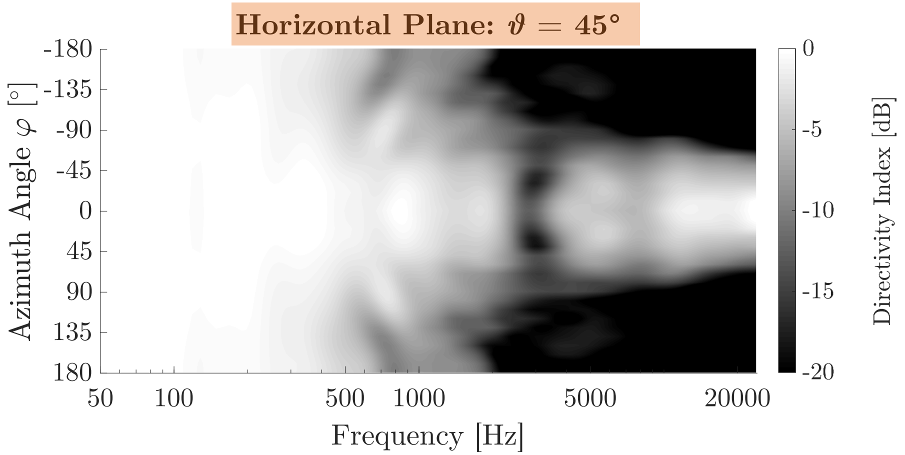
Example of a bassoon pattern

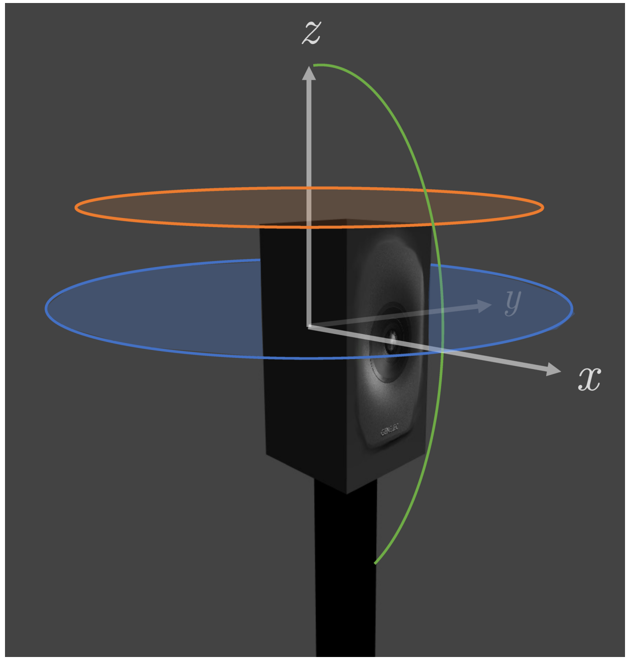


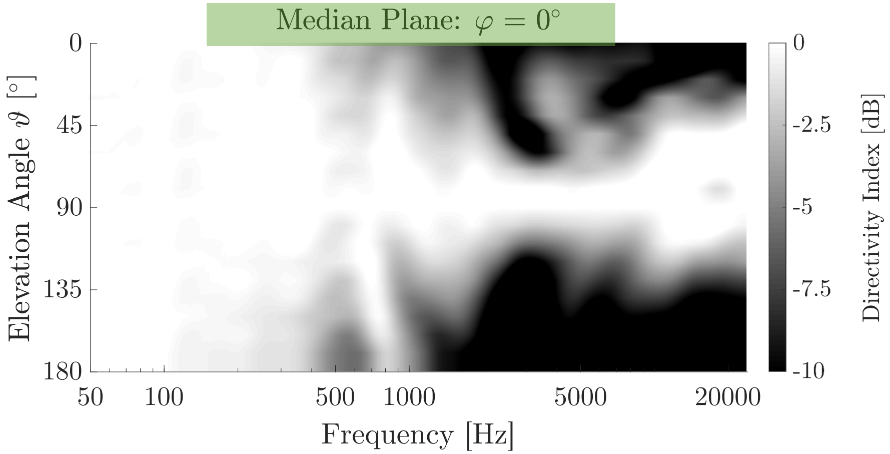
Ackermann, D., Brinkmann, F. & Weinzierl, S. A Database with Directivities of Musical Instruments. *J. Audio Eng. Soc.* **72**, 170–179 (2024).

Example of a loudspeaker directivity pattern







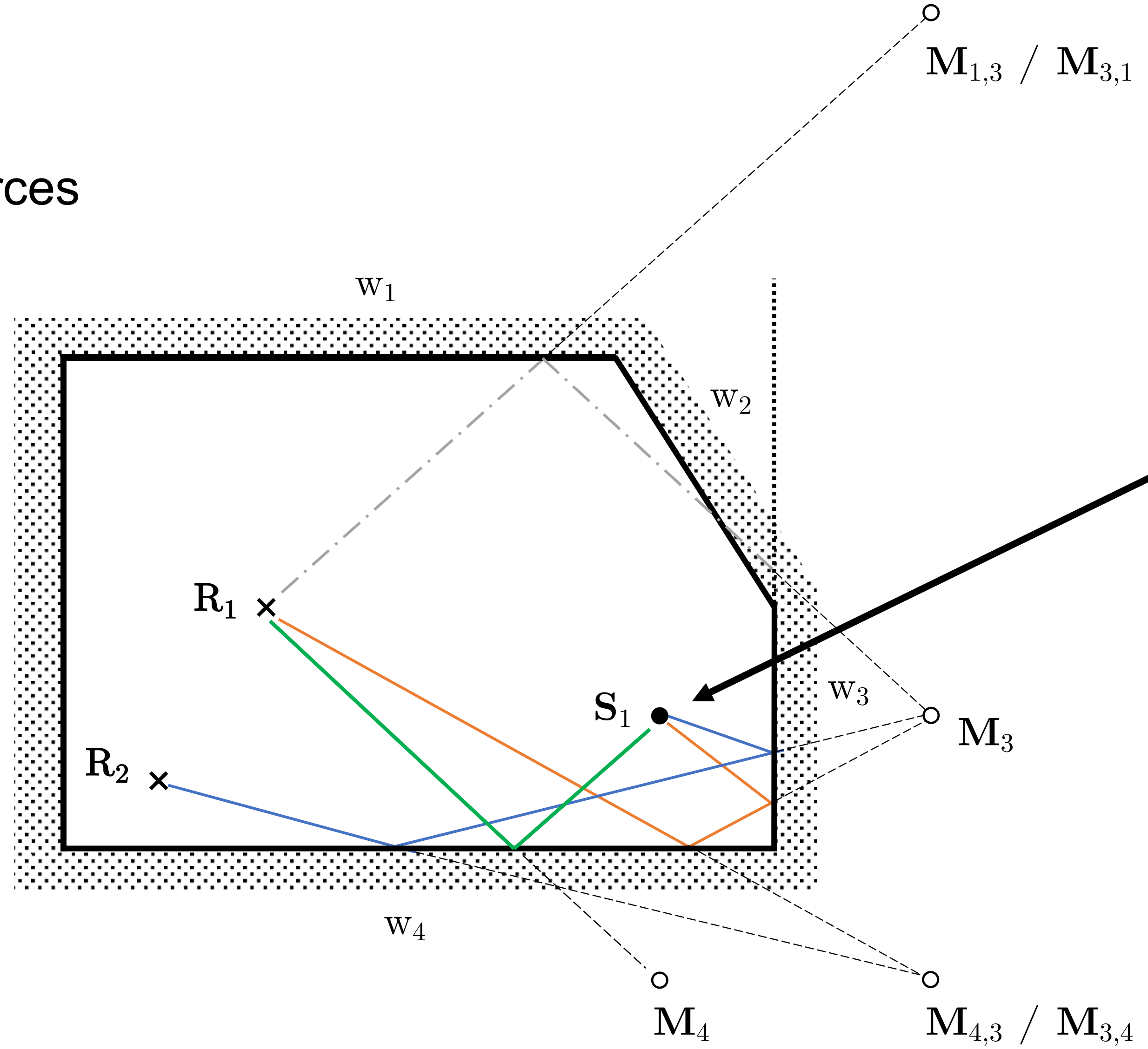


Götz, G. & Pulkki, V. Simplified Source Directivity Rendering in Acoustic Virtual Reality using the Directivity Sample Combination. in *147th Convention of the Audio Engineering Society* (New York, NY, USA, 2019).

A!

Source directivity in the image-source model

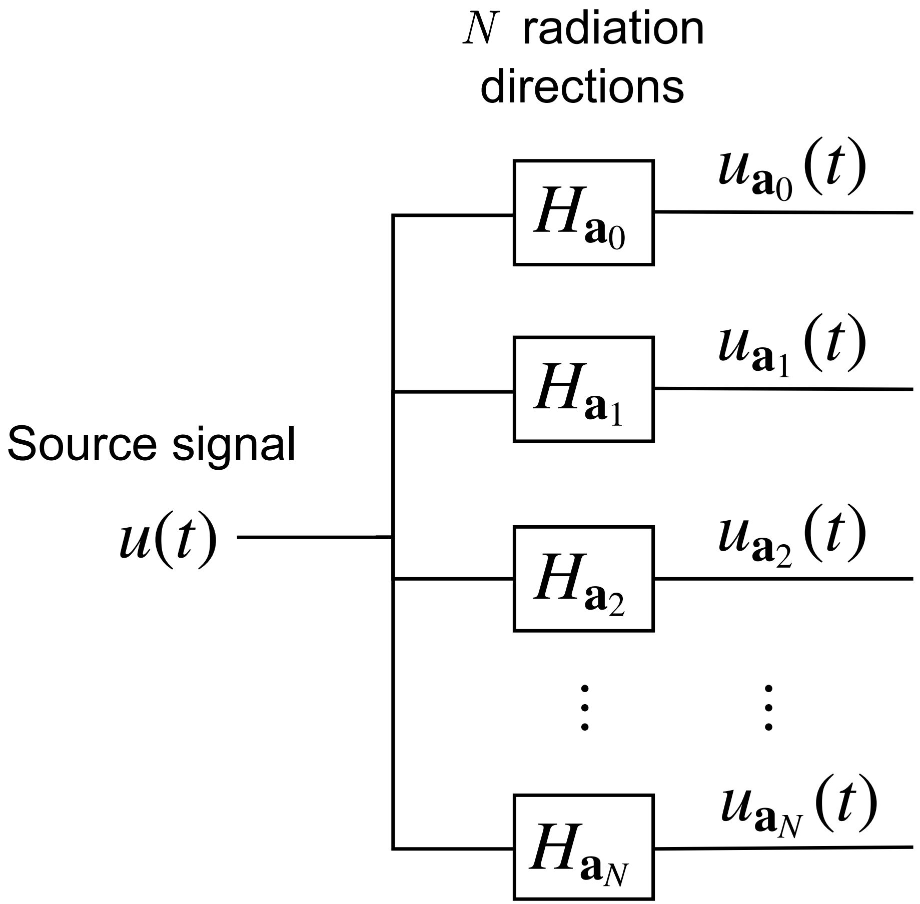
S: source
R: receivers
M: image-sources



Multiple radiation directions must be considered, even for a single source-receiver configuration!

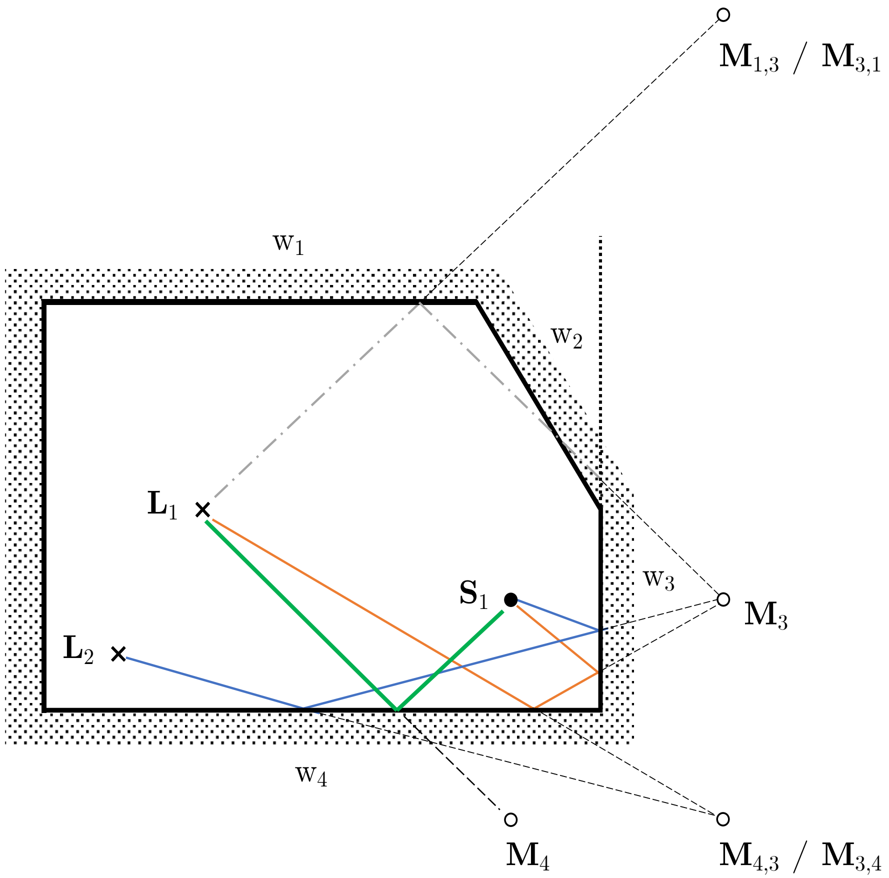
Source directivity can be included in audio engines by using filters for every radiation direction

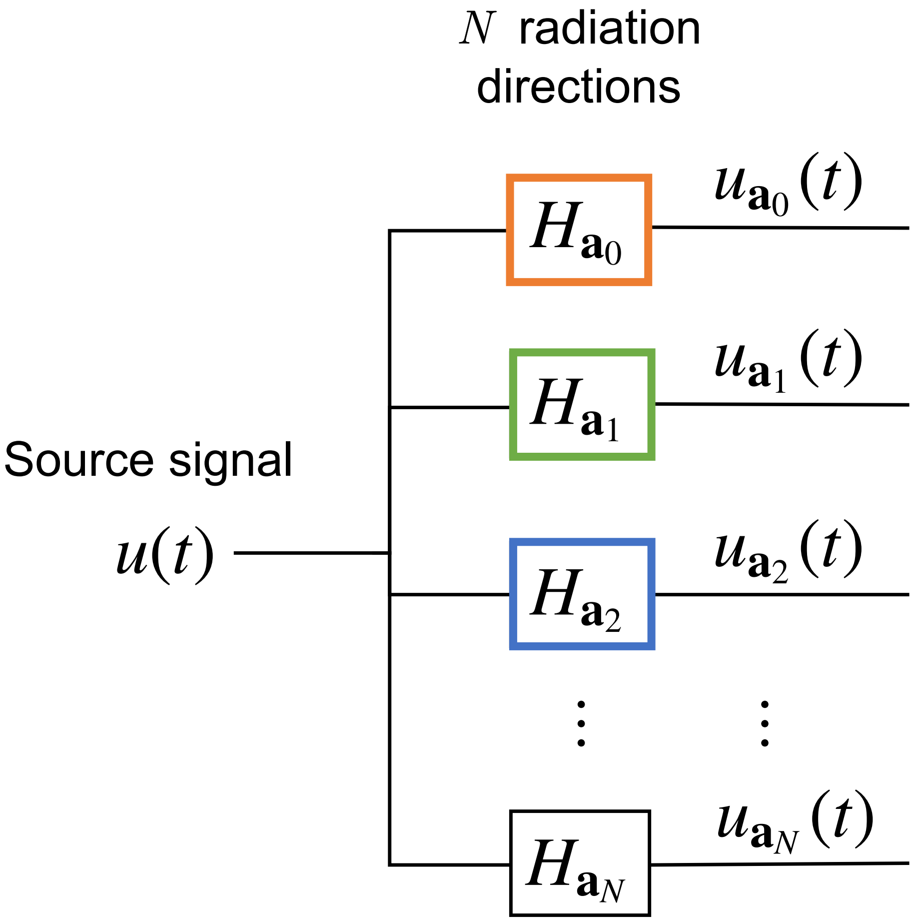
- Simulate multiple propagation paths, e.g. with image-source method or ray tracing
- Directivity filter for every radiation direction
  - FIR or IIR filters that fit the directivity pattern
  - Measured impulse responses
- For all propagation paths of all sources-receiver configurations in the scene



Savioja, L., Huopaniemi, J., Lokki, T. & Väänänen, R. Creating Interactive Virtual Acoustic Environments. *Journal of the Audio Engineering Society* **47**, 675–705 (1999).

Source directivity can be included in audio engines by using filters for every radiation direction





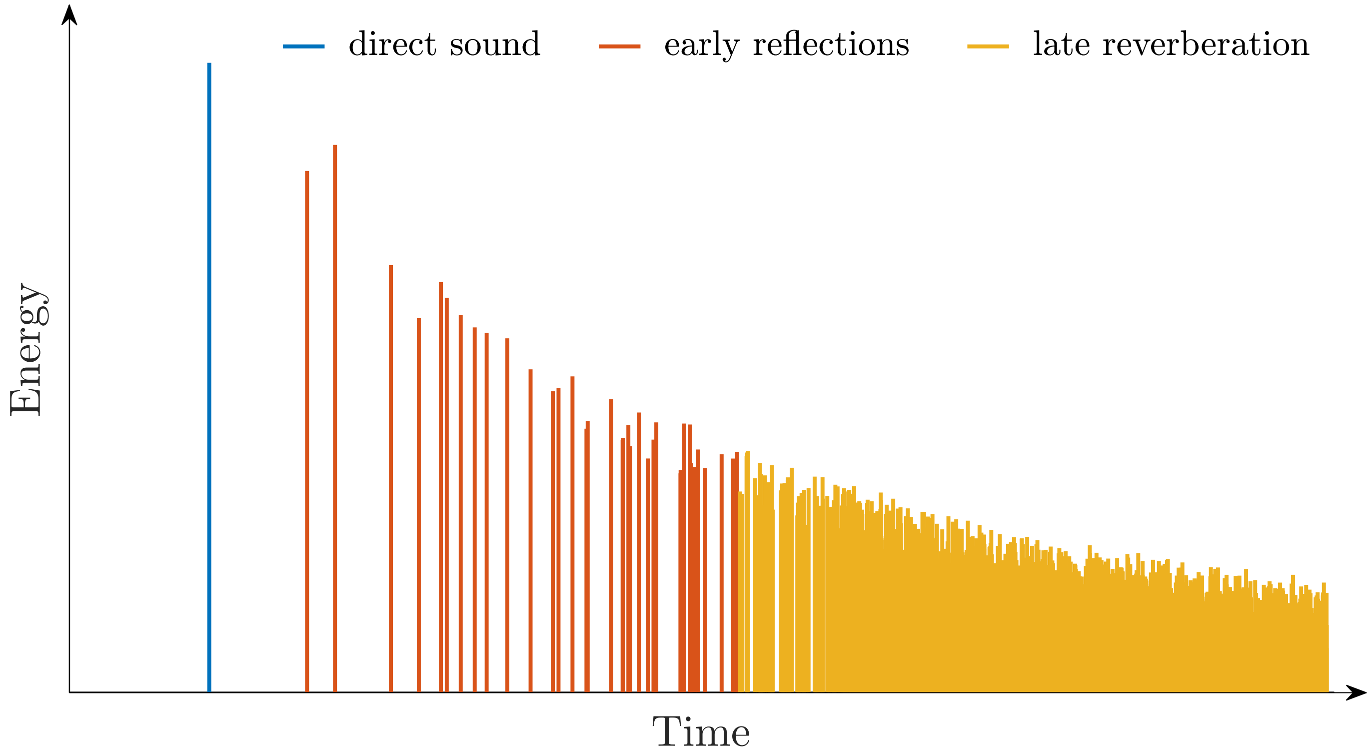
A!

Savioja, L., Huopaniemi, J., Lokki, T. & Väänänen, R. Creating Interactive Virtual Acoustic Environments. *Journal of the Audio Engineering Society* **47**, 675–705 (1999).

Room-acoustic modelling

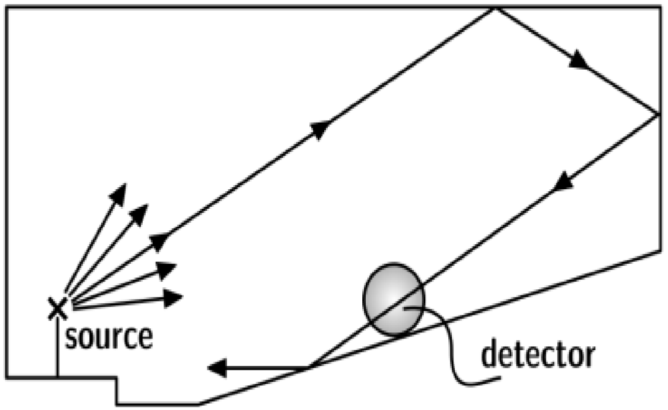
Sound propagation between a source and a receiver can be described using room impulse responses

Squared room impulse response



A!

Room-acoustic modelling has traditionally been of interest in architectural planning and design; various approaches exist



Ray tracing

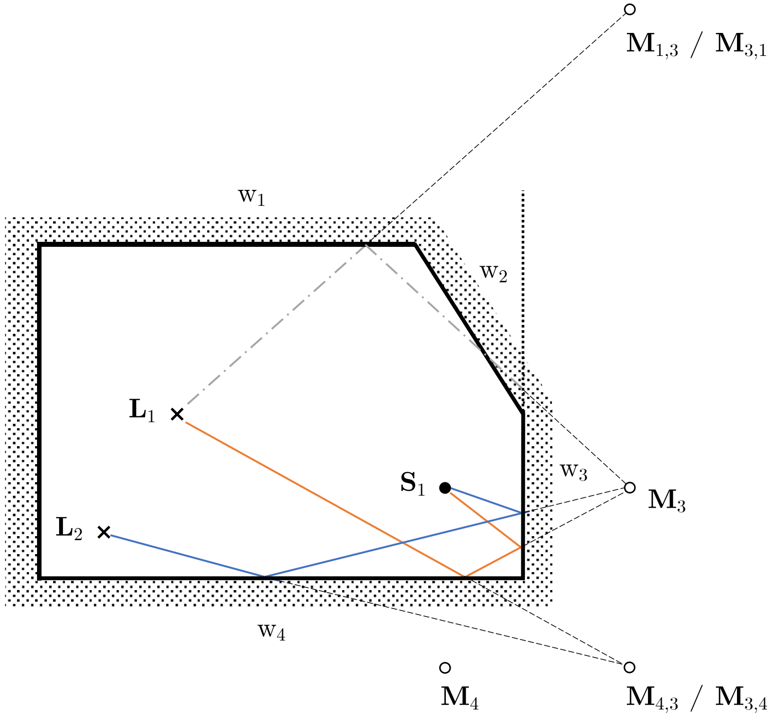
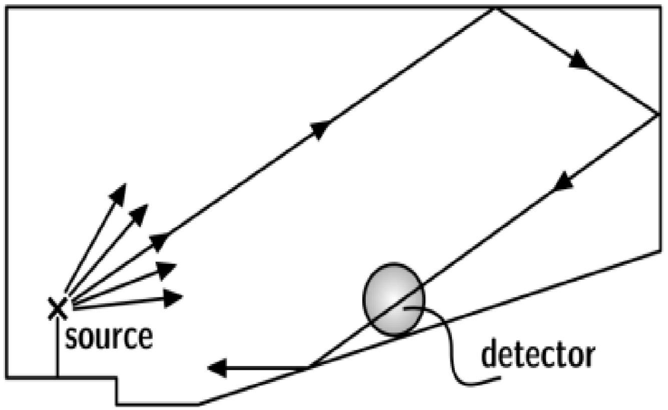


Image-source method

Wave-based simulations

Vorländer, M. *Auralization*. (Springer, 2008)

Room-acoustic modelling has traditionally been of interest in architectural planning and design



Ray tracing

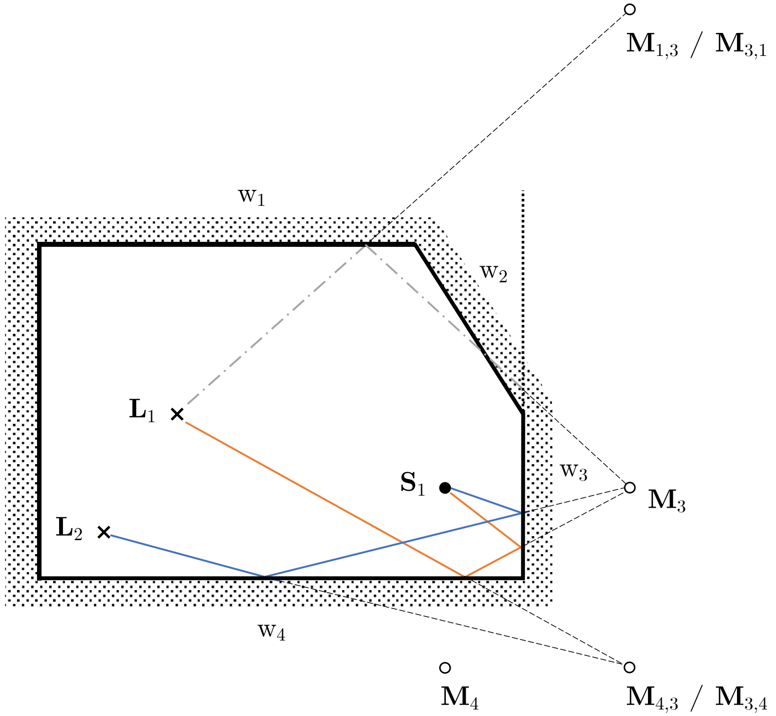
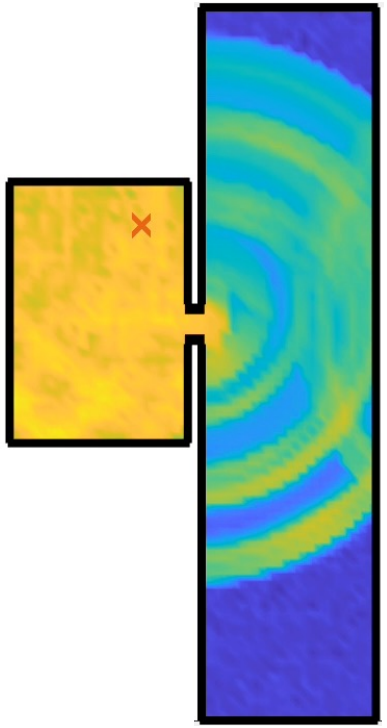


Image-source method

Geometrical acoustics



Wave-based simulations

Savioja, L. & Svensson, U. P. Overview of geometrical room acoustic modeling techniques. *J Acoust Soc Am* **138**, 708–730 (2015).

Bilbao, S. *Wave and Scattering Methods for Numerical Simulation*. (John Wiley & Sons, 2004).

A!

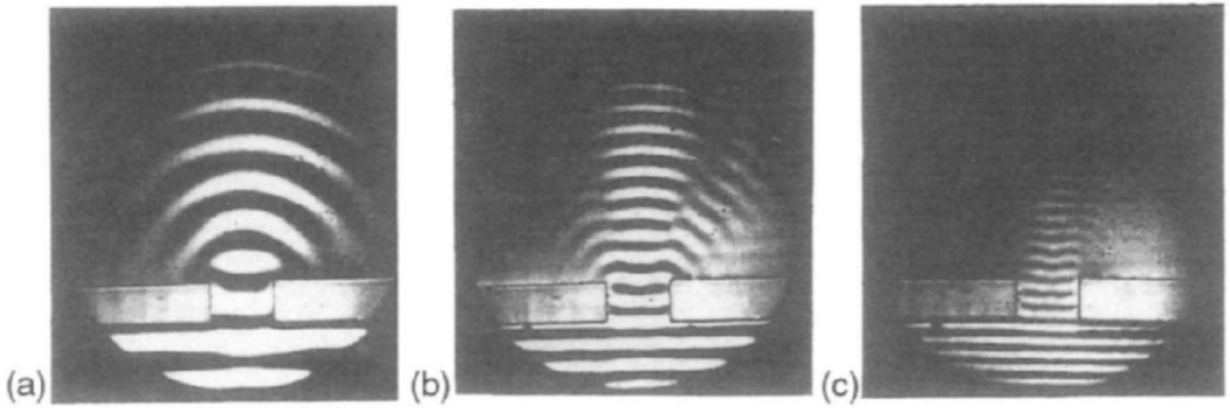
Both simulation paradigms have strengths and drawbacks

Geometrical acoustics:

+ Fast (real-time)
+ Easy to implement for simple geometries (e.g. shoebox)

- Difficult to model wave effects, such as diffraction. (Diffraction can be important near doors and when transitioning between rooms!)

Diffraction at aperture, e.g. door



Low frequency | Medium frequency | High frequency

Fahy, F. *Foundations of Engineering Acoustics*. (2000).

Both simulation paradigms have strengths and drawbacks

Geometrical acoustics:

+ Fast (real-time)
+ Easy to implement for simple geometries (e.g. shoebox)

- Difficult to model wave effects, such as diffraction. (Diffraction can be important near doors and when transitioning between rooms!)

Wave-based simulations:

+ Physically accurate (solving the wave-equation)

- Slow and computationally expensive (typically not real-time)

Both paradigms rely heavily on the quality of the input data (3D mesh and material properties).
Garbage in → Garbage out

Room-acoustic simulation requires different inputs

Room geometry

3D mesh

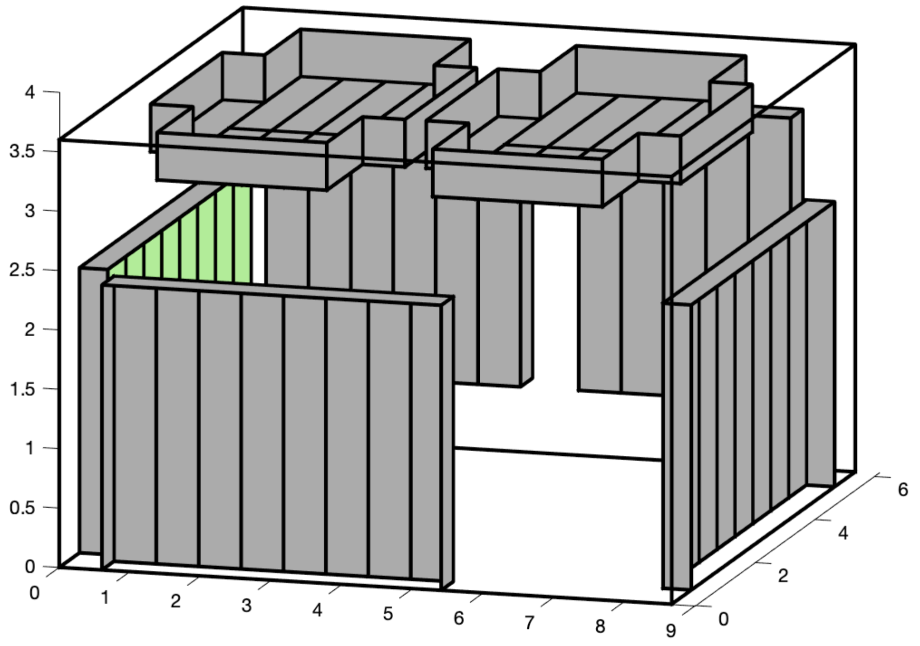
Acoustic material properties

- Must be available for each surface
- Absorption + scattering coefficients

Source-receiver configuration

Required for calculating

- Propagation delays
- Attenuation due to air absorption
- Attenuation due to wall absorption on the propagation path



State-of-the-art game engines

Wwise

Core features

- Simple source directivity
- Distance attenuation
- Dynamic early reflections
- Late reverberation
- Diffraction, occlusion
- Portals
- 3D spatialization

Reverberation simulation and rendering

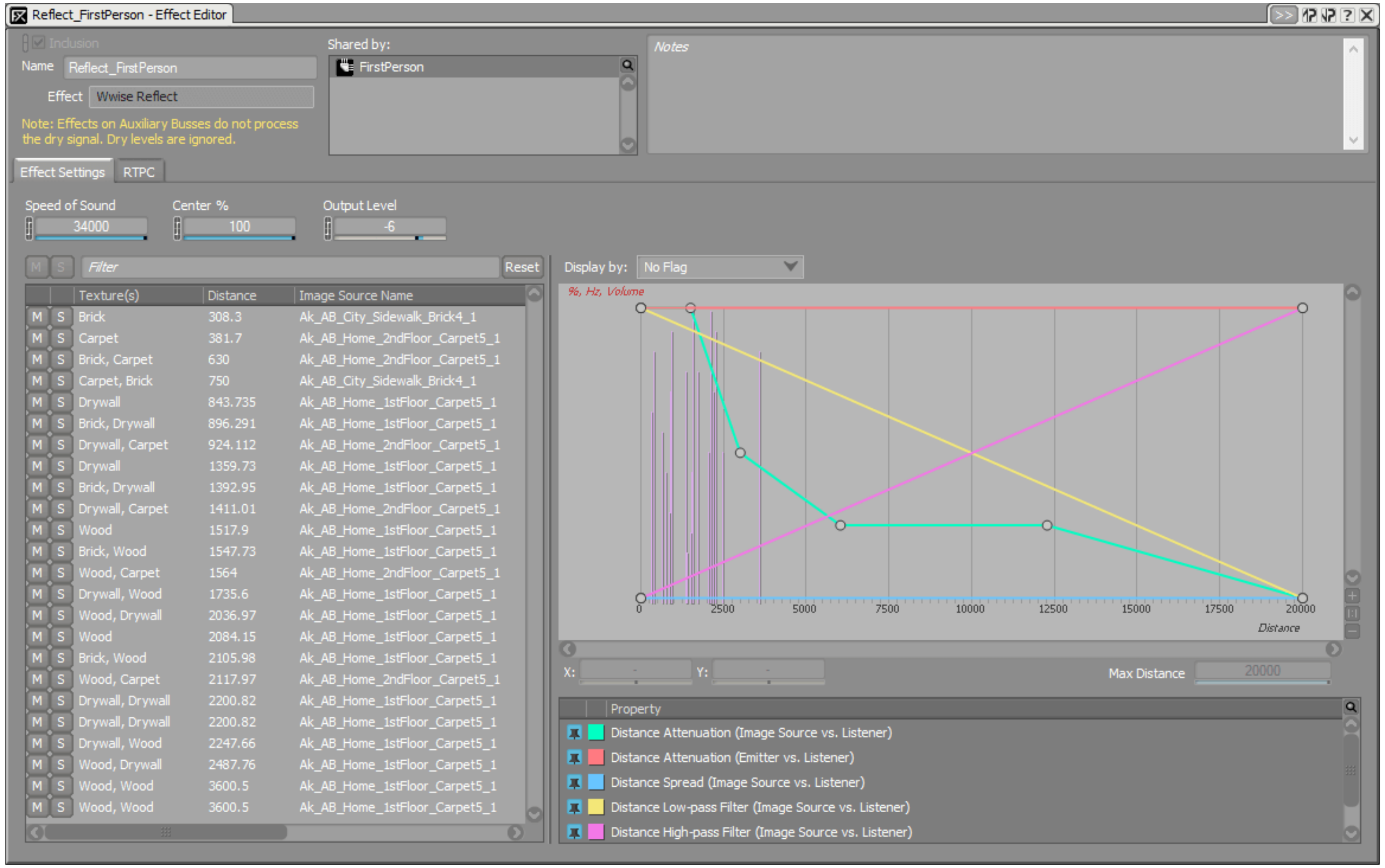
- Convolution with room impulse responses
- Early reflections rendering using image-source method
- Reverberators with artistic parameters

Integration into game engines

- Unity
- Unreal
- ….

A!

Wwise "Reflect" plugin can dynamically update early reflections from the image-source method

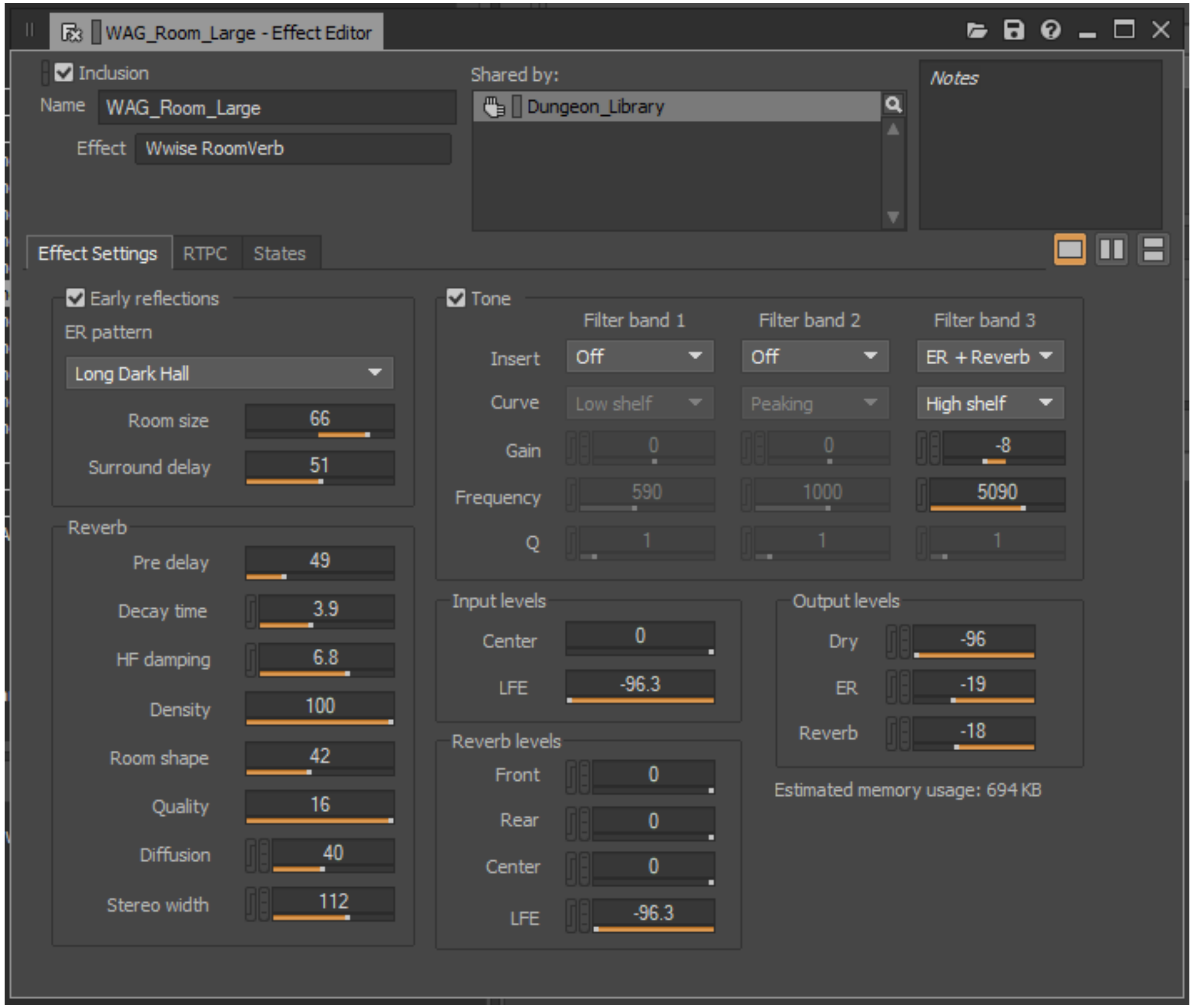


- Up to 4th order
- Incl. distance attenuation
- Incl. diffraction and obstruction

A!

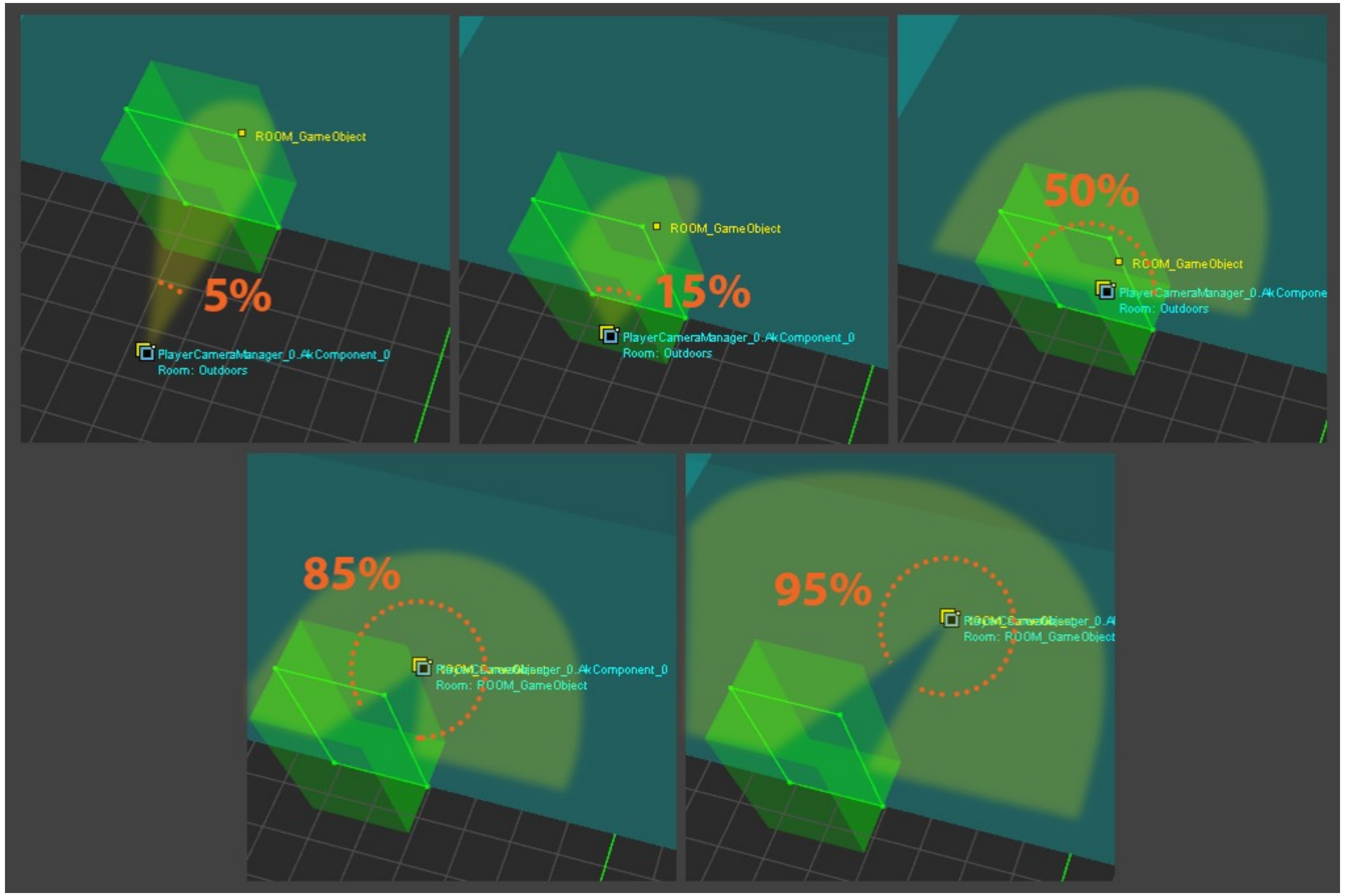
https://www.audiokinetic.com/en/wwise/plugins/reflect/

Wwise “RoomVerb” plugin can be used to reverberate sounds based on artistic parameters



https://www.audiokinetic.com/en/courses/wwise251/?source=wwise251&id=Lesson4_Wwise_RoomVerb#read

Wwise uses portals to model sound transmission between two rooms, e.g. through doors or windows



"Portal spread" controls how much sound spills into the other room

A!

https://blog.audiokinetic.com/en/rooms-and-portals-with-wwise-spatial-audio/

Steam audio

Core features
- Simple source directivity
- Sound propagation
- Occlusion
- Realtime updates
- 3D spatialization
- Baked propagation

Reverberation simulation and rendering
- Ray tracing

Integration into game engines
- Unity
- Unreal
- ….

https://valvesoftware.github.io/steam-audio/doc/unity/guide.html

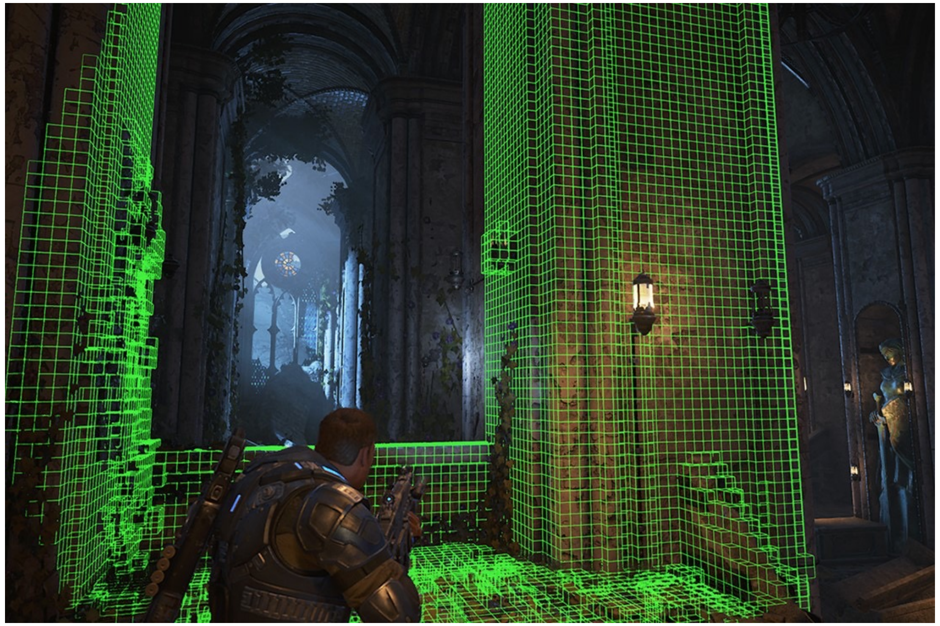
Microsoft Project Acoustics / Triton

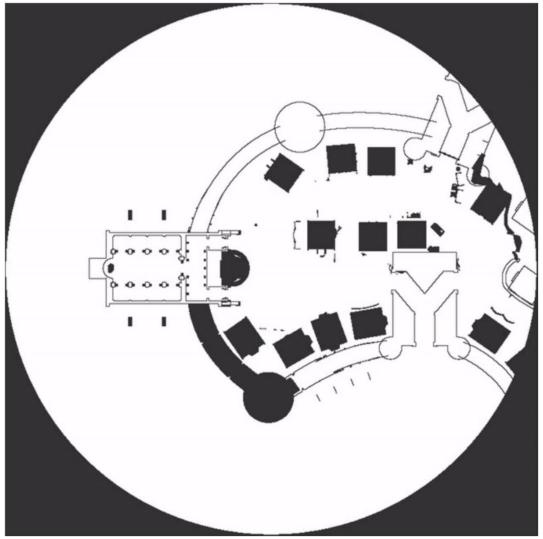
Core features

- Wave-based simulation: before game; in cloud
- Baking results and render them during runtime
- Directional sources and receivers

Integration into game engines

- Unity
- Unreal
- ….



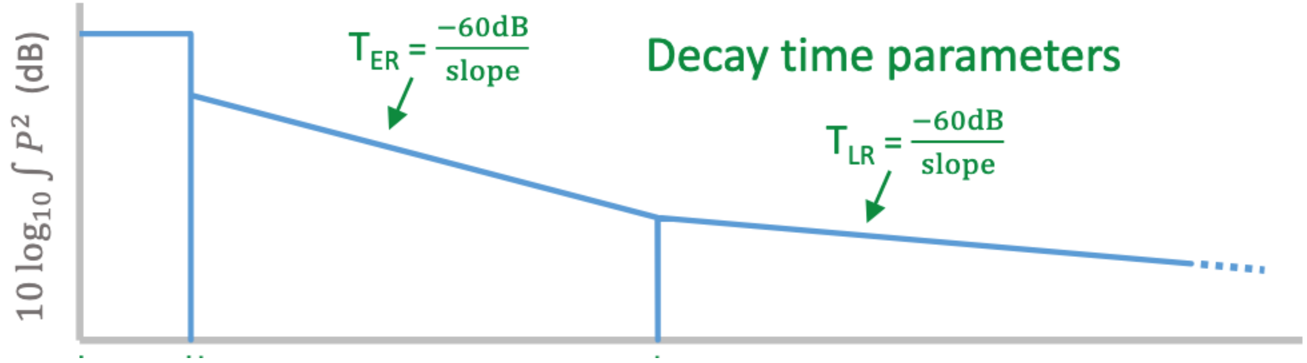


https://learn.microsoft.com/en-us/gaming/acoustics/what-is-acoustics

Project Acoustics extracts various parameters from wave-based simulations for real-time rendering

Direct sound and early reflections
- Time delay
- Level/Loudness
- Direction

Late reverberation
- Decay times
- Levels/Loudness



Raghuvanshi, N. & Snyder, J. Parametric wave field coding for precomputed sound propagation. *ACM Transactions Graph* **33**, 38 (2014)

Raghuvanshi, N. & Snyder, J. Parametric directional coding for precomputed sound propagation. *ACM Transactions Graph* **37**, 108 (2018).

Chaitanya, C. R. A. *et al.* Directional sources and listeners in interactive sound propagation using reciprocal wave field coding. *ACM Trans. Graph.* **39**, 44:1-44:14 (2020).

A!

The parameters are available at different probe locations

https://learn.microsoft.com/en-us/gaming/acoustics/unreal-baking-probes

Project Acoustics: demo

https://youtu.be/qCUEGvIgco8?feature=shared&t=645

Demo can be seen at 10:45

Aalto-yliopisto
Aalto-universitetet
Aalto University

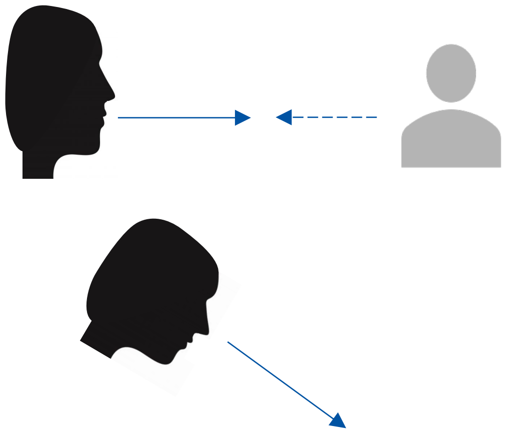
Research at the Aalto Acoustics Lab towards heuristic and data-driven game audio engines

Aalto-yliopisto
Aalto-universitetet
Aalto University

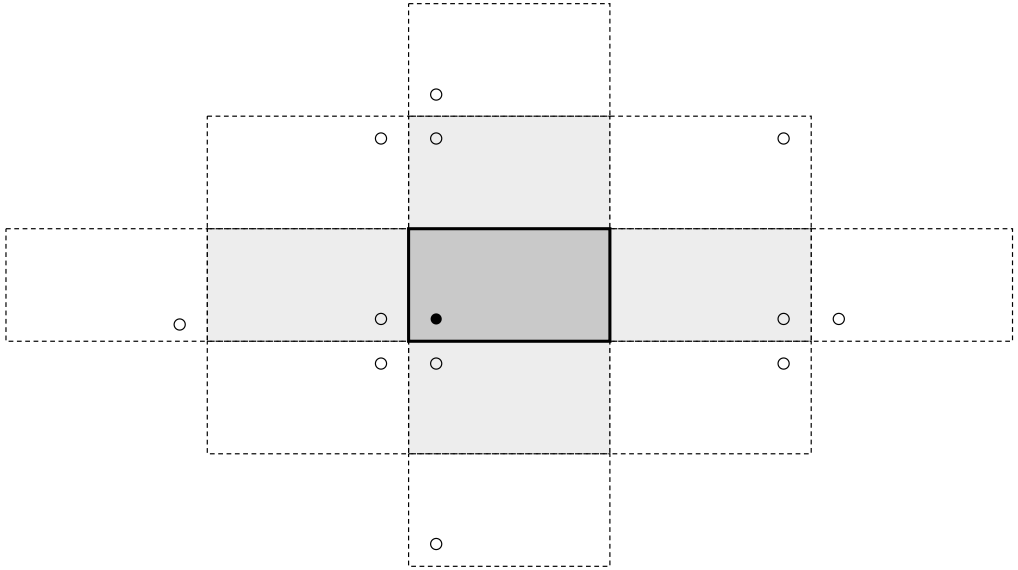
1. Directivity rendering using the Directivity Sample Combination (DISCO)

Götz, G. & Pulkki, V. Simplified Source Directivity Rendering in Acoustic Virtual Reality using the Directivity Sample Combination. in *147th Convention of the Audio Engineering Society* (New York, NY, USA, 2019).

Number of directivity filters must be reduced for real-time rendering in game audio engines



Simultaneous auralization of many sound sources



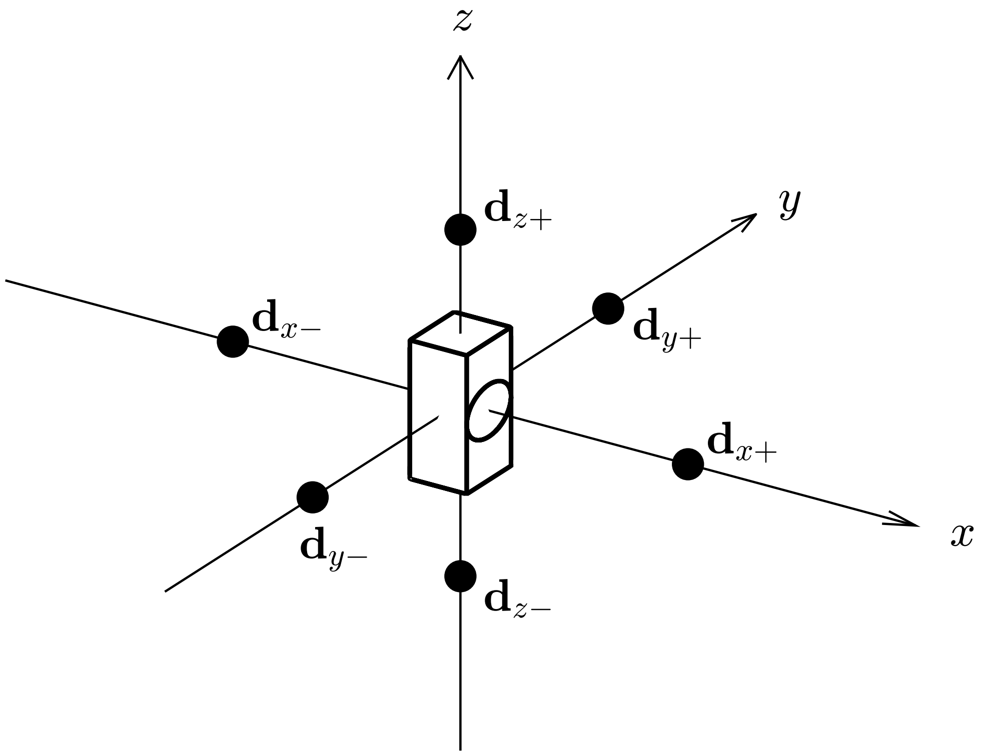
Early reflection rendering with image-sources for complicated geometries and/or high reflection orders

A!

Sampling the directivity pattern in certain directions captures characteristic acoustic properties

Interpolation for intermediate directions

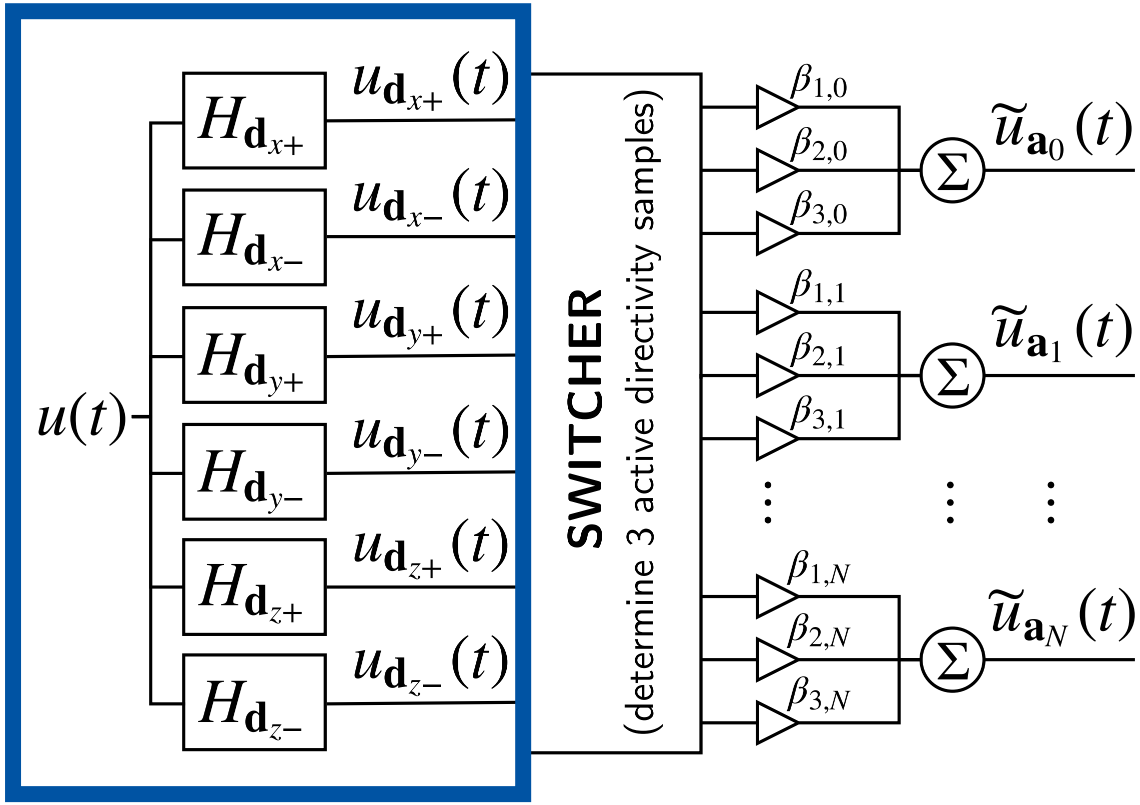
In our paper: 6 directivity samples at axis directions



The proposed implementation requires only 6 directivity filters

Interpolation: VBAP-inspired formulation

- Only 6 directivity filters

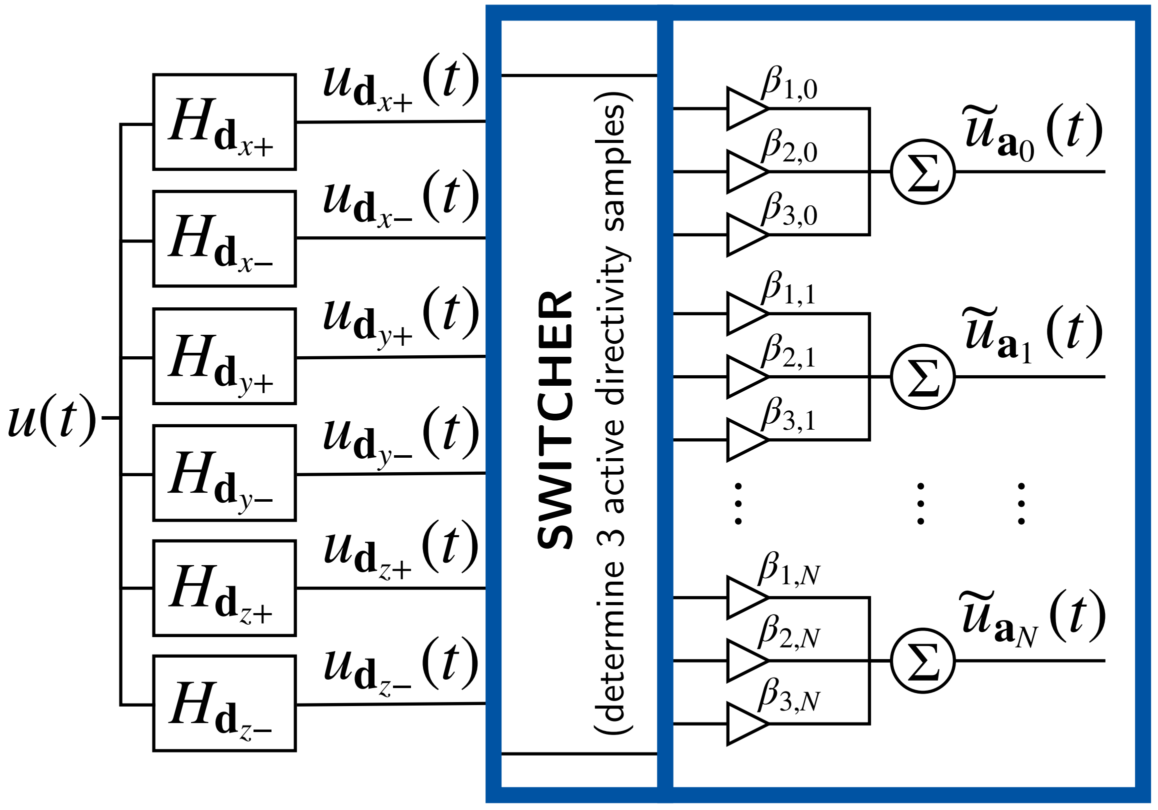


A!

The proposed implementation requires only 6 directivity filters

Interpolation: VBAP-inspired formulation

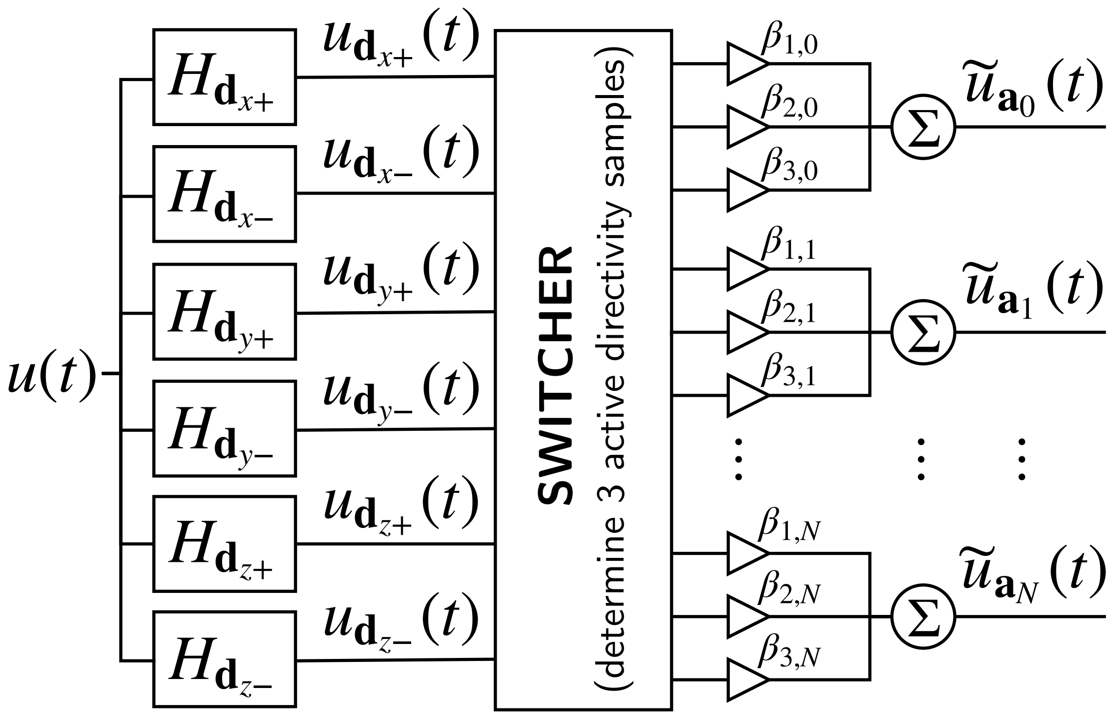
- Only 6 directivity filters
- Determine 3 active directivity samples
- Weight and sum the corresponding filter outputs



The proposed implementation requires only 6 directivity filters

Interpolation: VBAP-inspired formulation

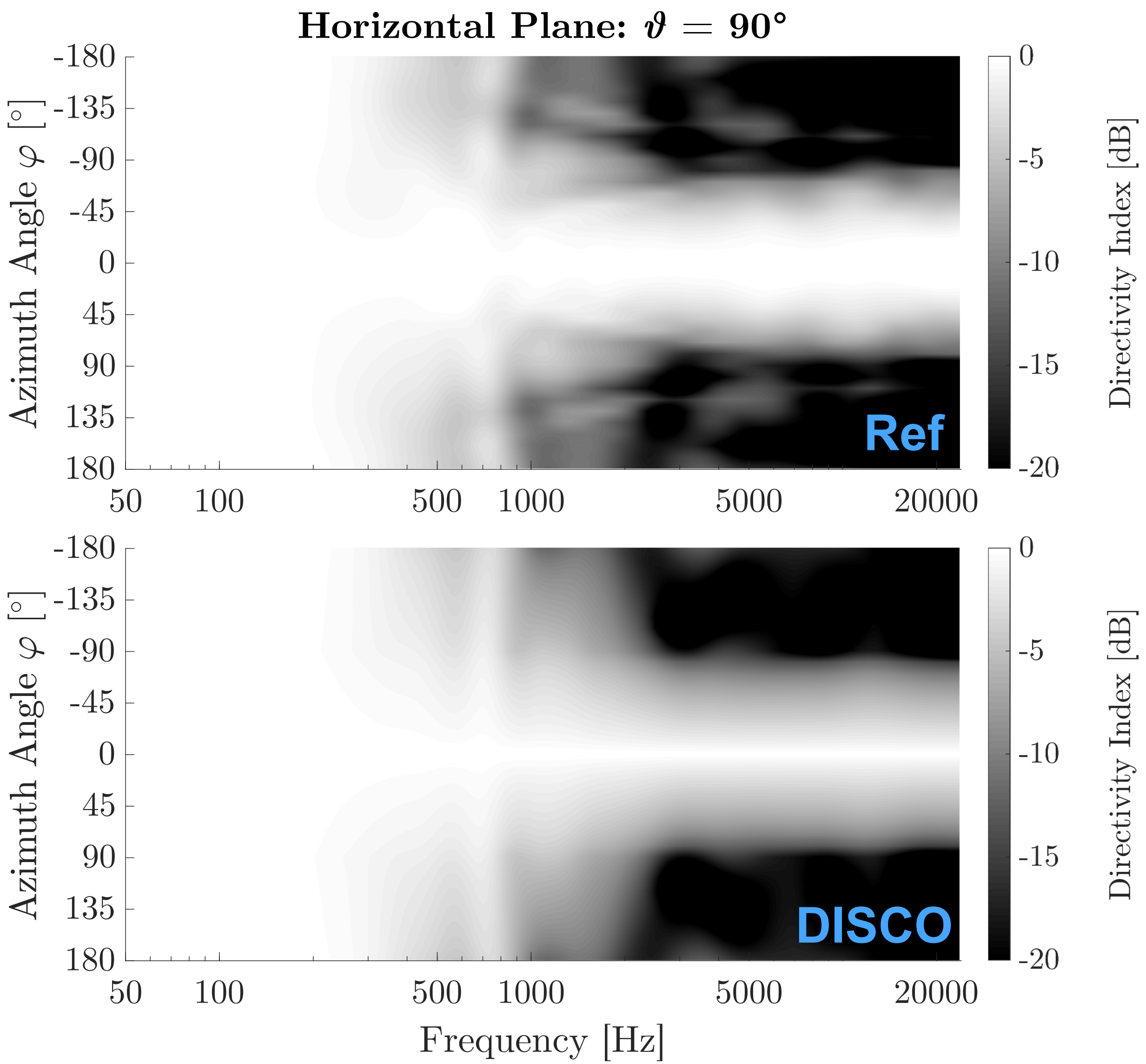
- Only 6 directivity filters
- Determine 3 active directivity samples
- Weight and sum the corresponding filter outputs



DISCO preserves frequency dependency

Loudspeaker directivity pattern (Ref):
Genelec 8020

Application of the DISCO simplification to the pattern

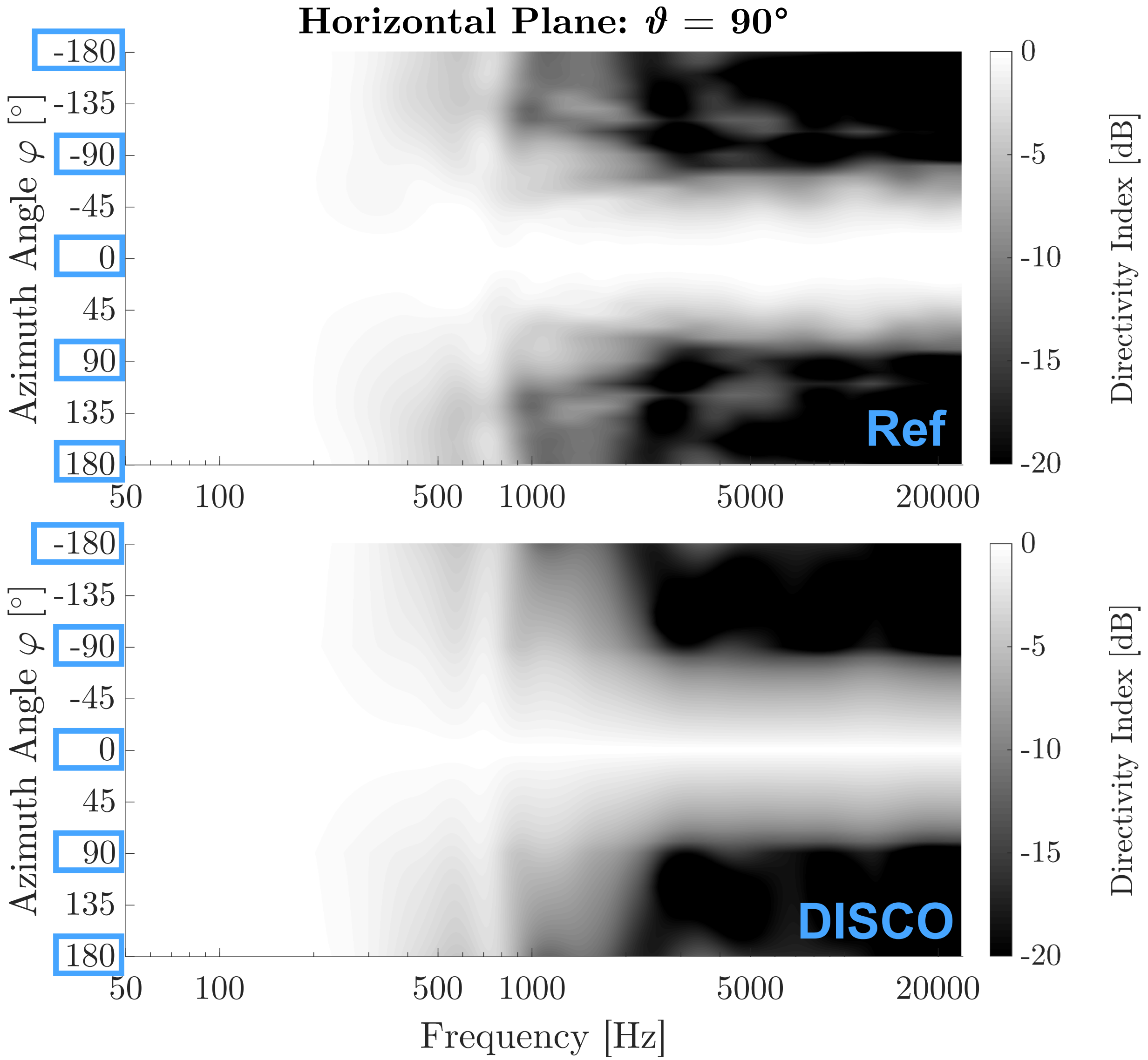


A!

DISCO preserves frequency dependency

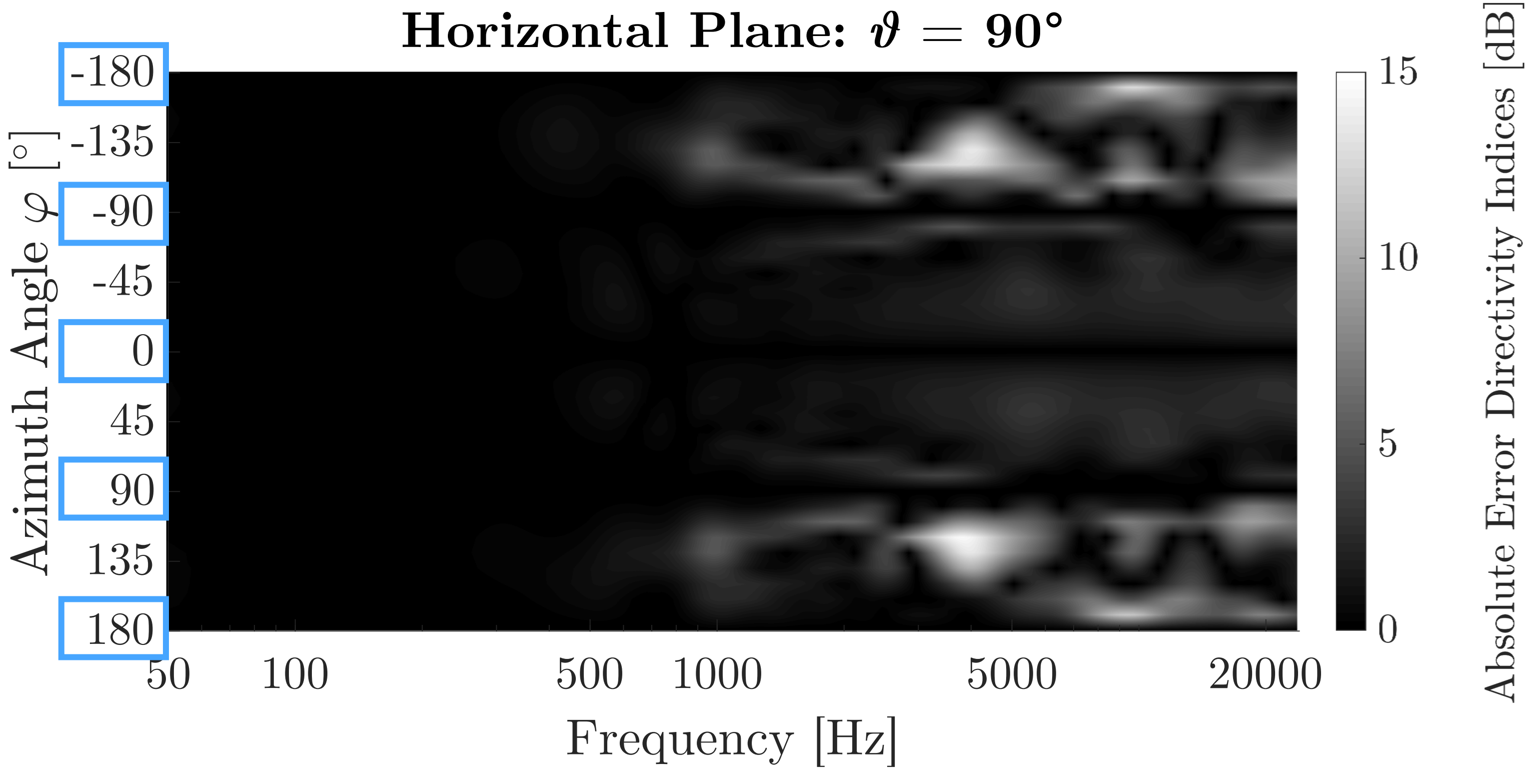
Loudspeaker directivity pattern (Ref):
Genelec 8020

Application of the DISCO simplification to the pattern (directivity samples in every axis direction)



A!

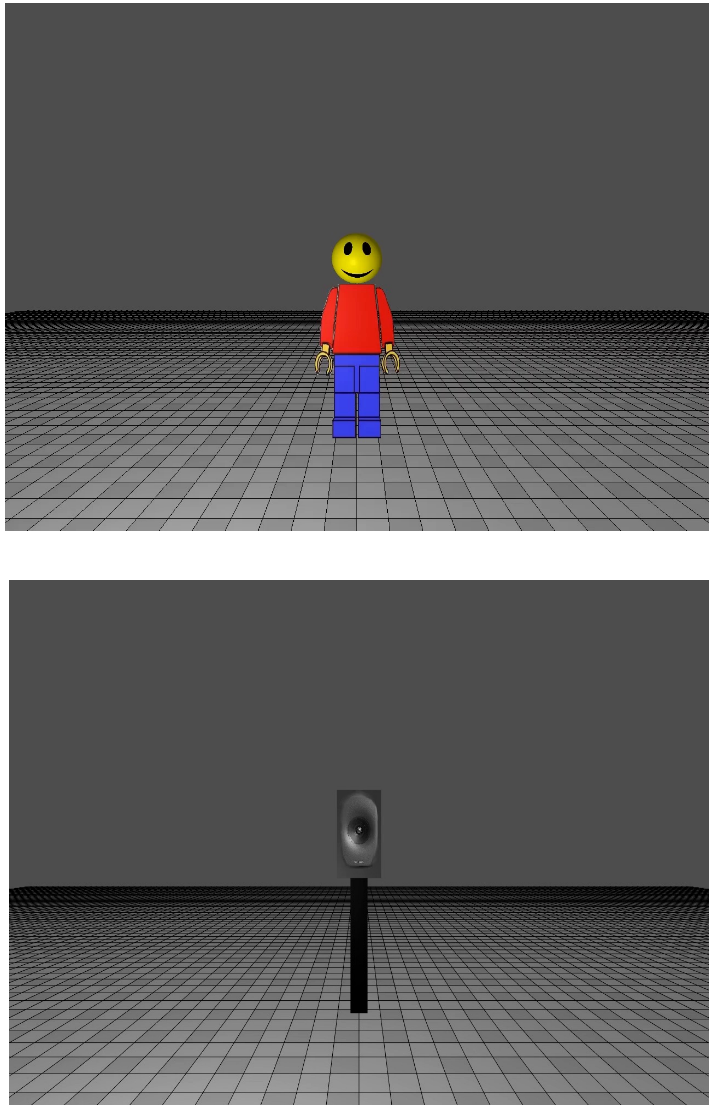
Absolute error between reference pattern and DISCO simplification reveals loss of directional details



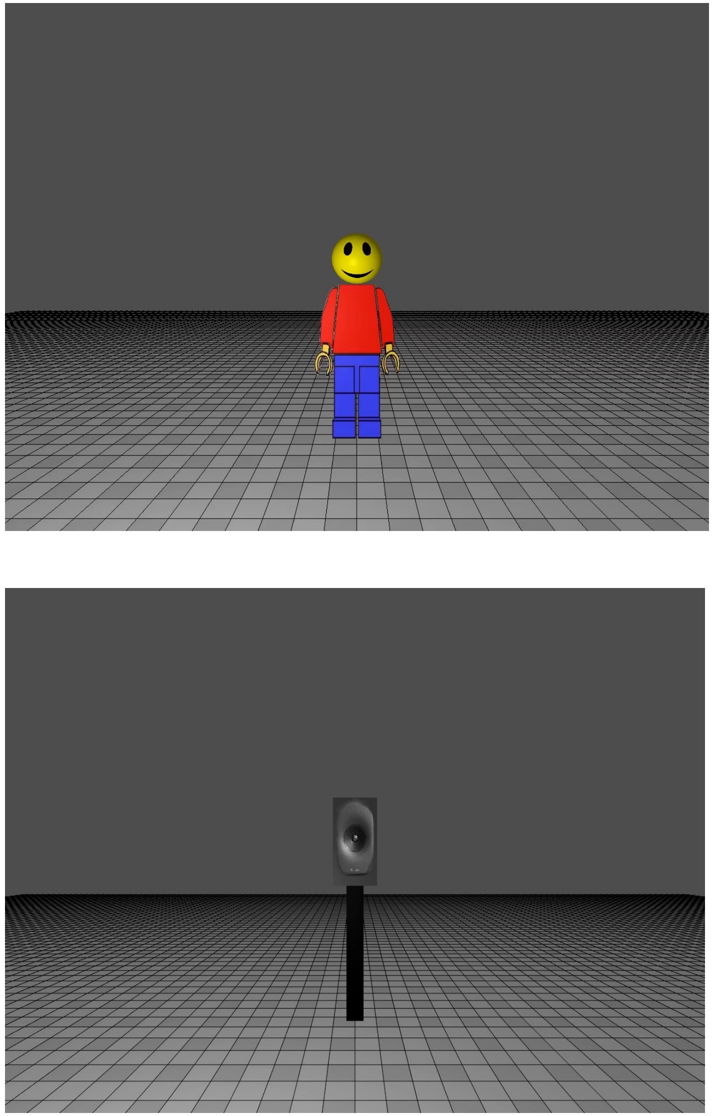
$$\varepsilon(\vartheta, \varphi) = |D_{\mathrm{ref}}(\vartheta, \varphi) - D_{\mathrm{DISCO}}(\vartheta, \varphi)|$$

A!

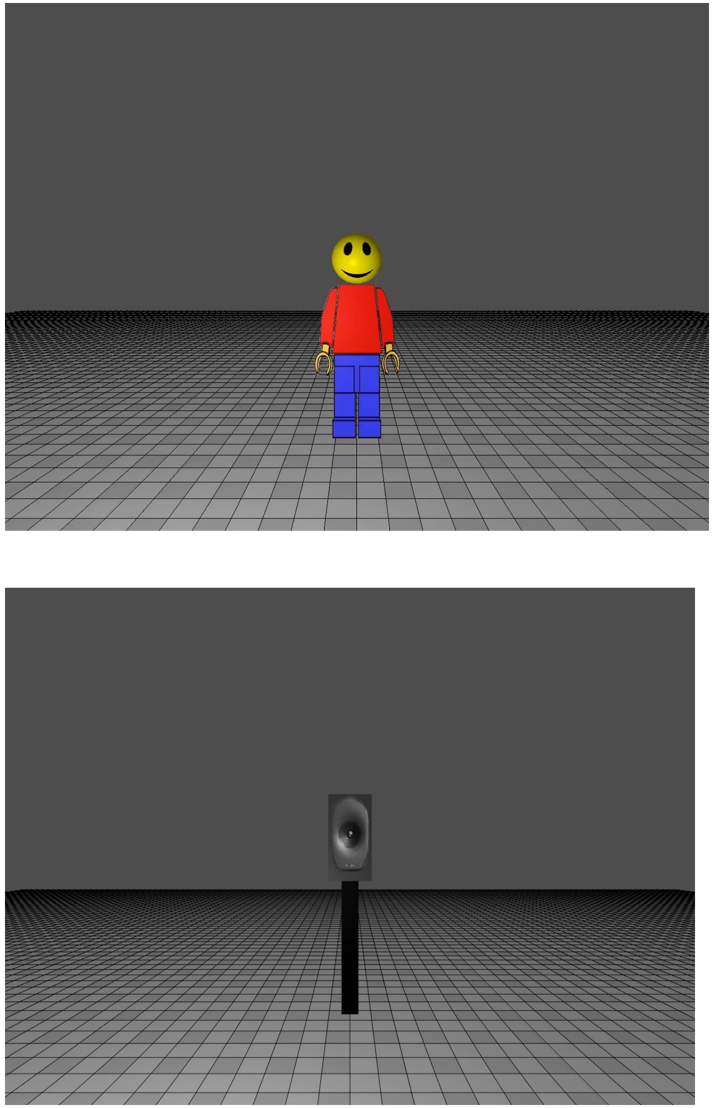
Demo



Reference



DISCO



Cardioid

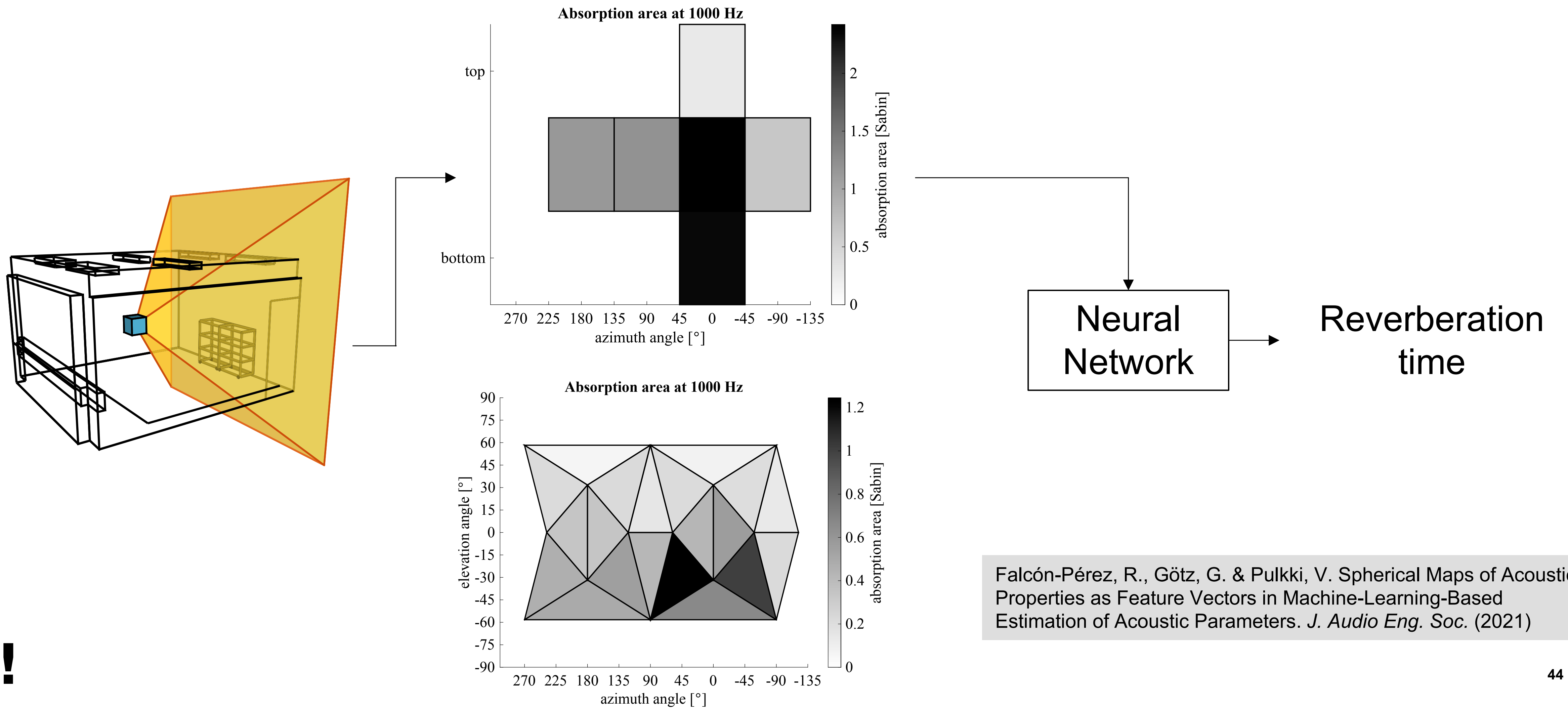
2. Dynamic late reverberation rendering using the common-slope model

G. Götz, T. Kerimovs, S. J. Schlecht, and V. Pulkki. Dynamic late reverberation rendering using the common- slope model. Accepted for publication in *AES 6th International Conference on Audio for Games*, Tokyo, Japan, April 2024.

How to use parametric rendering for late reverberation in an audio engine?

Can we avoid the pre-computation step?

We started by predicting reverberation time from projections of absorption areas

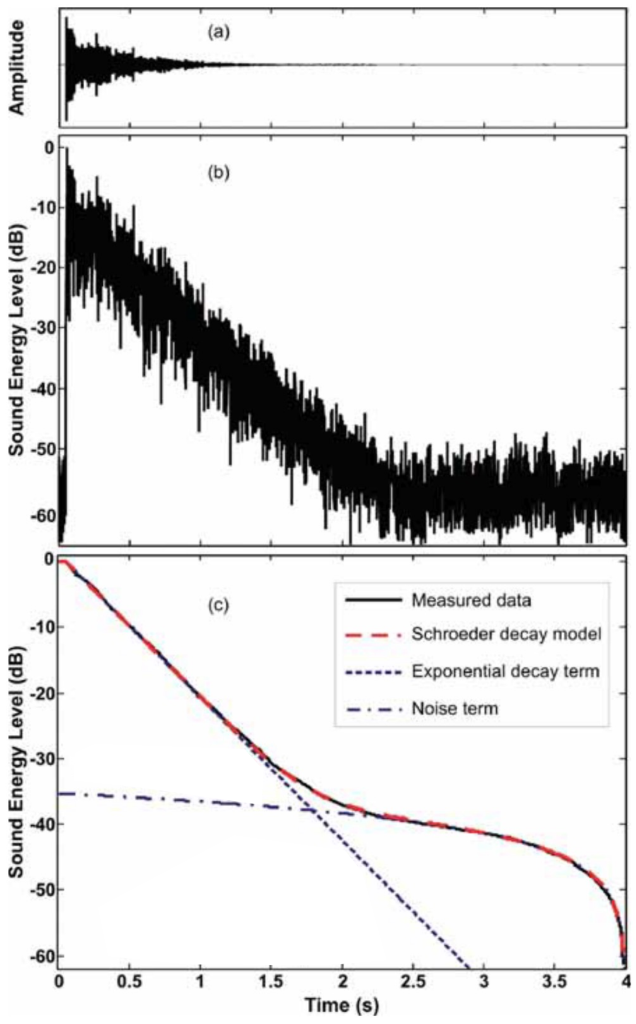


Falcón-Pérez, R., Götz, G. & Pulkki, V. Spherical Maps of Acoustic Properties as Feature Vectors in Machine-Learning-Based Estimation of Acoustic Parameters. *J. Audio Eng. Soc.* (2021)

How to model late reverberation in dynamic and more general environments?

A!

Energy decay functions can be used to describe late reverberation



A

Room impulse response (*RIR*)

$$h(\mathbf{x}, t)$$

Energy-time function (*squared RIR*)

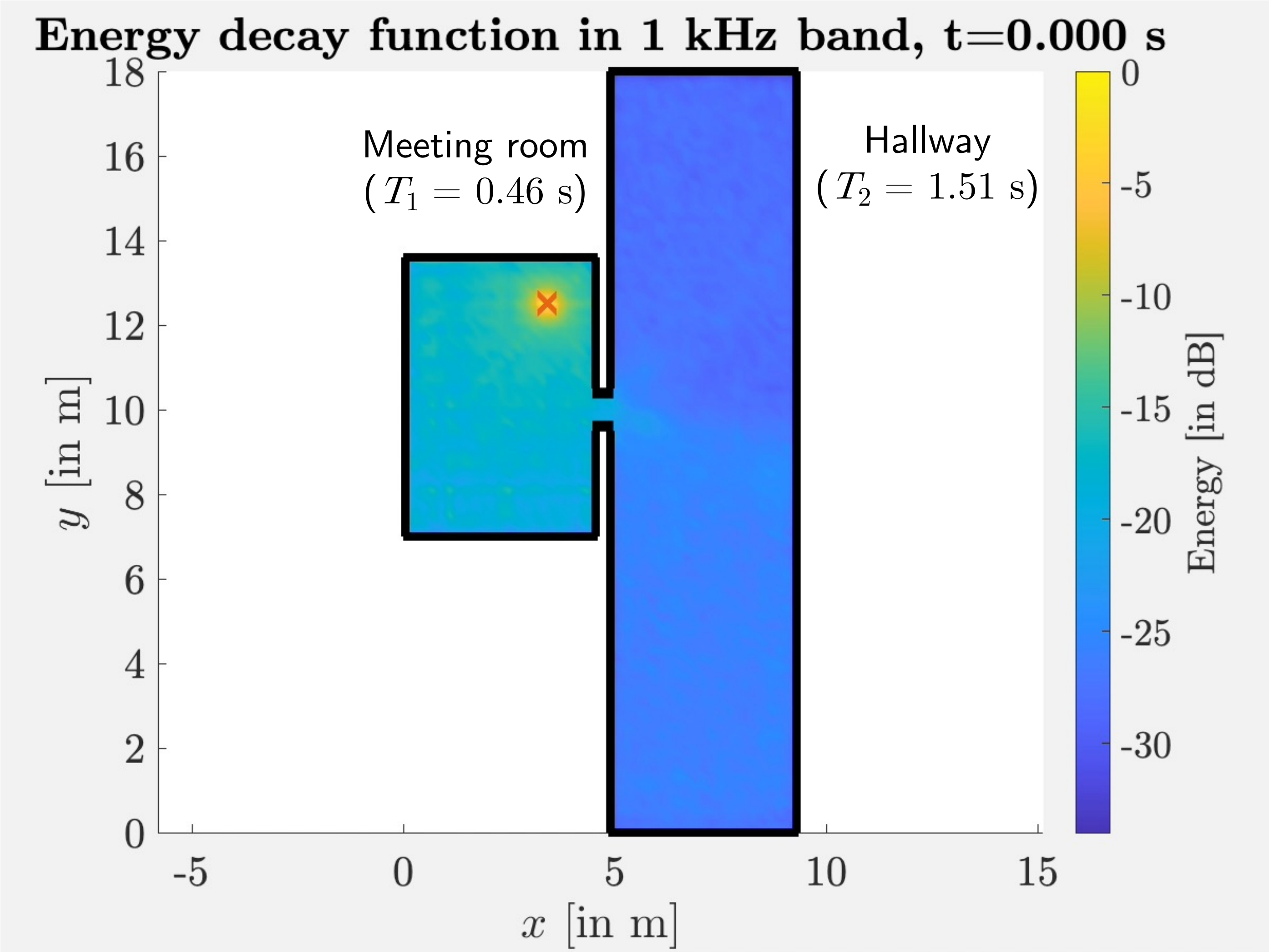
$$h^2(\mathbf{x}, t)$$

Energy decay function (*backwards-integrated squared RIR*)

$$d(\mathbf{x}, t) = \sum_{l=t}^{L} h^2(\mathbf{x}, l)$$

Xiang, N. *Architectural Acoustics Handbook*. (J. Ross Publishing, 2017)

Inhomogeneity: late reverberation varies spatially



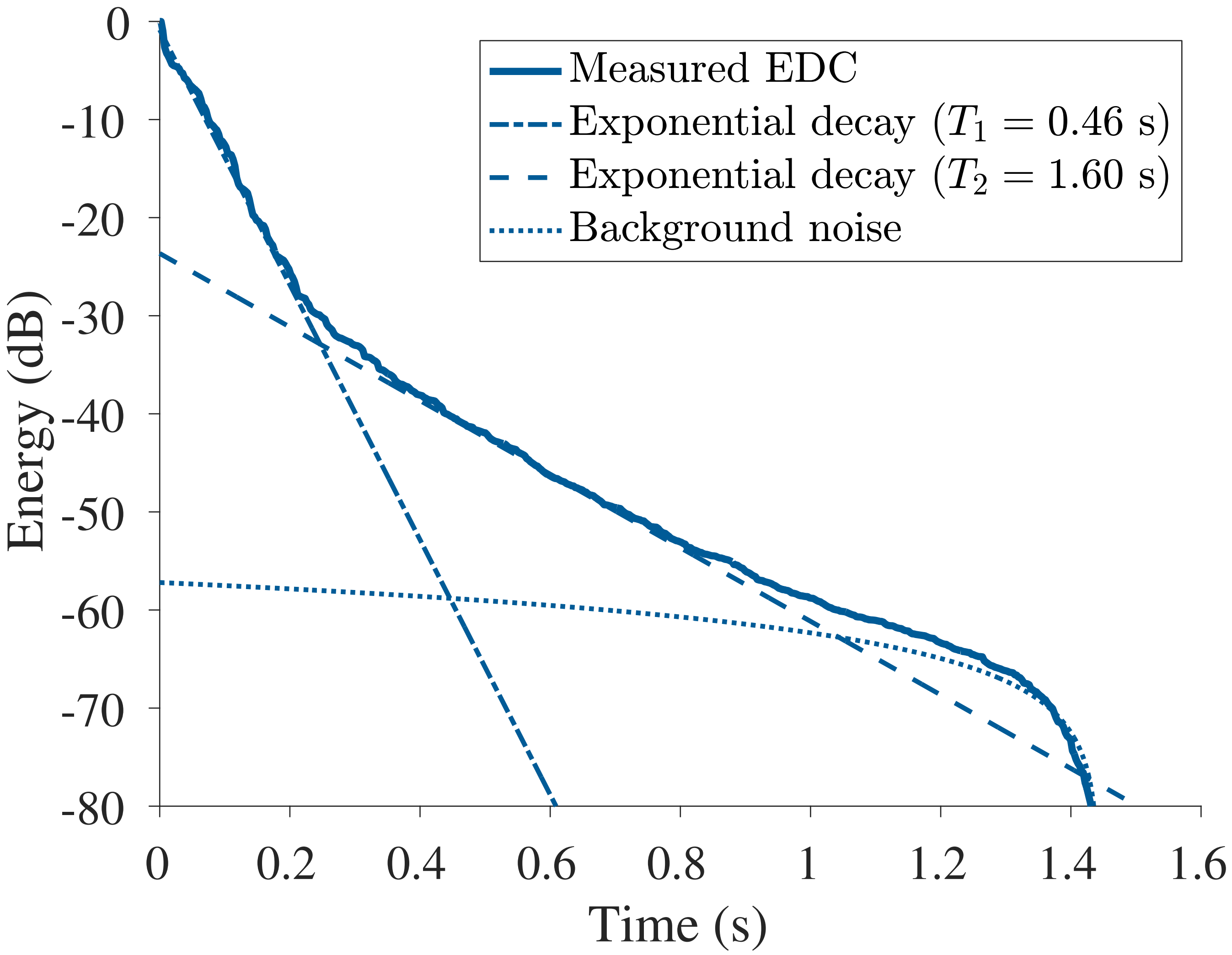
... and also directionally (not shown here)

The animation shows how energy decays inside a scene with two coupled rooms

1. Before t=0, we had a sound source playing at x, and it is turned off at t=0
2. Meeting room is less reverberant → energy gets absorbed quicker
3. Hallway is more reverberant → energy does not get absorbed as quickly

A!

Sound energy decay can be modelled as a superposition of multiple decaying exponentials



Determine model parameters from EDC, for example using a neural network:

Götz, G., Pérez, R. F., Schlecht, S. J. & Pulkki, V. Neural network for multi-exponential sound energy decay analysis. *J Acoust Soc Am* **152**, 942–953 (2022).

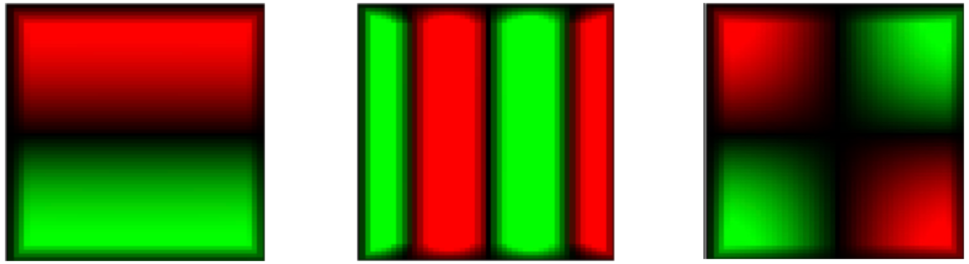
A!

Mode decay times are independent of the source-receiver configuration

Modal decomposition of RIR:

$$h(\mathbf{x}, t) = \sum_{m=1}^{M} \chi_m(\mathbf{x}) \; \tau_m(t)$$

Spatial term $\chi_m(\mathbf{x})$



Temporal term $\tau_m(t)$

$$\tau_m(t) = \cos\left(\frac{\omega_m t}{f_\mathrm{s}}\right) \exp\left(-\frac{\delta_m t}{f_\mathrm{s}}\right)$$

does not depend on $\mathbf{x}$

Kuttruff, H. *Room Acoustics*. (Spon Press, 2000)

Haneda, Y., Kaneda, Y. & Kitawaki, N. Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function. *IEEE Trans. Speech Audio Process. P* **7**, 709–717 (1999).

Slope decay times are also independent of the source-receiver configuration

“Traditional” multi-exponential model

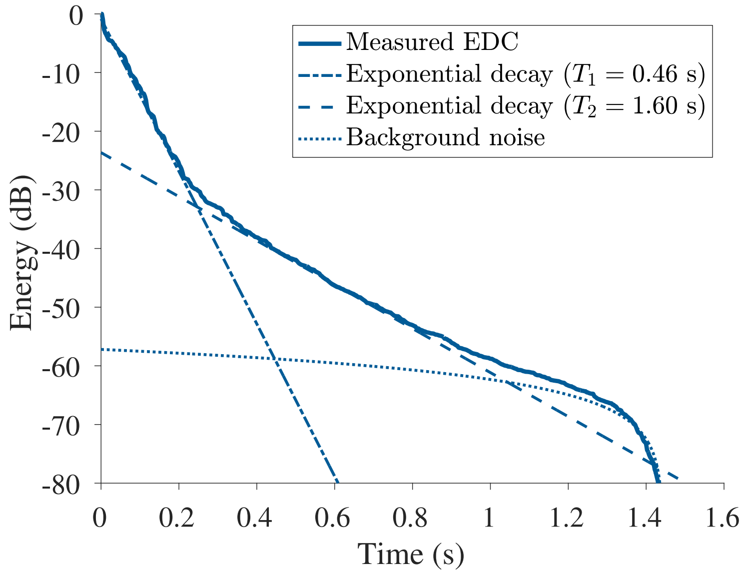
$$d_K^{(\mathrm{tr.})}(\mathbf{x},t) = N_{0,\mathbf{x}} \ \Psi_{0,\mathbf{x}}^{(\mathrm{tr.})}(t) \ + \sum_{k=1}^{K} A_{k,\mathbf{x}} \left[\Psi_{k,\mathbf{x}}^{(\mathrm{tr.})}(t) - \Psi_{k,\mathbf{x}}^{(\mathrm{tr.})}(L)\right]$$

$$\Psi_{k,\mathbf{x}}^{(\mathrm{tr.})}(t) = \begin{cases} L-t\,, & \text{if } k=0 \\ \exp\left(\frac{-13.8}{f_s}\frac{t}{T_{k,\mathbf{x}}}\right), & \text{if } k>0 \end{cases}.$$

Common-slope model

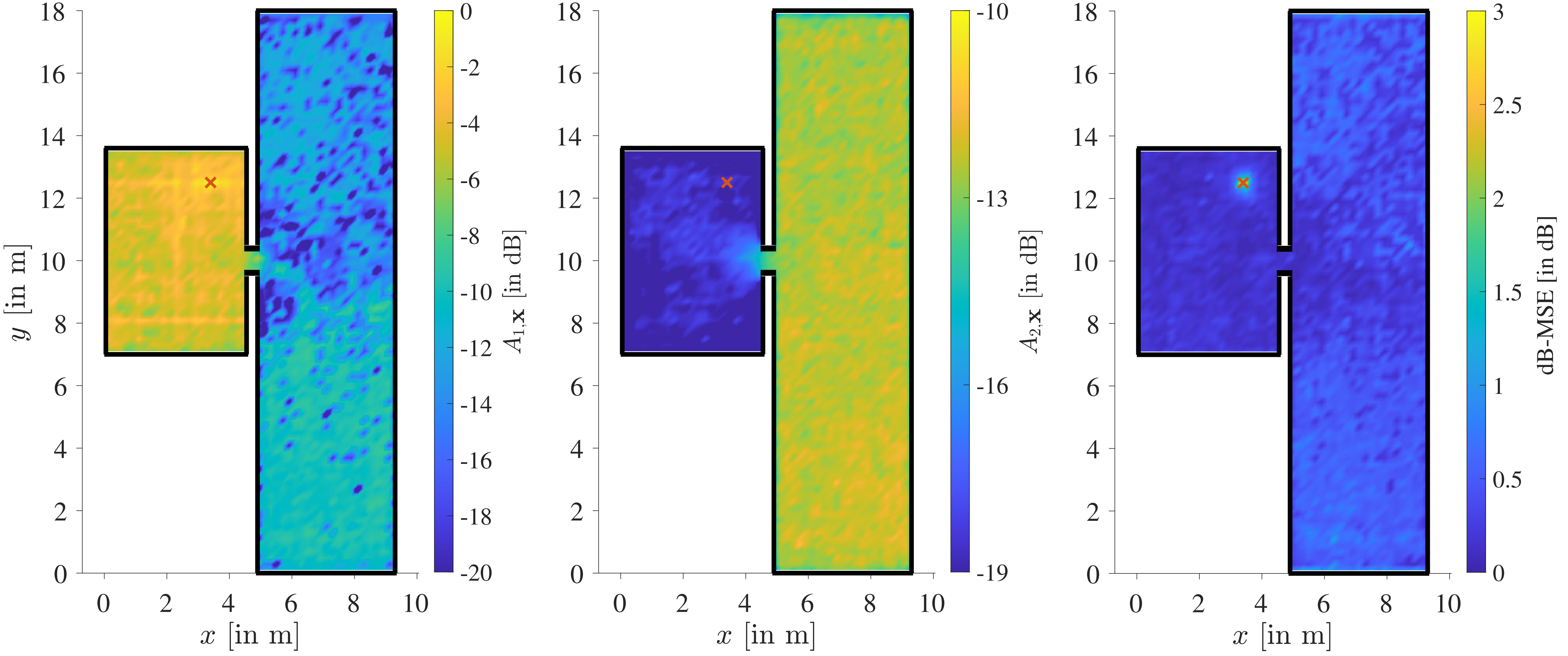
$$d_K(\mathbf{x},t) = N_{0,\mathbf{x}} \ \Psi_0(t) \ + \sum_{k=1}^{K} A_{k,\mathbf{x}} \left[\Psi_k(t) - \Psi_k(L)\right],$$

$$\Psi_k(t) = \begin{cases} L-t\,, & \text{if } k=0 \\ \exp\left(\frac{-13.8}{f_s}\frac{t}{T_k}\right), & \text{if } k>0 \end{cases}.$$



Götz, G., Schlecht, S. J. & Pulkki, V. Common-slope modeling of late reverberation. *IEEEACM Trans. Audio, Speech, Lang. Process.* **31**, 3945–3957 (2023).

Common-slope analysis illustrates reverberation fade in coupled rooms

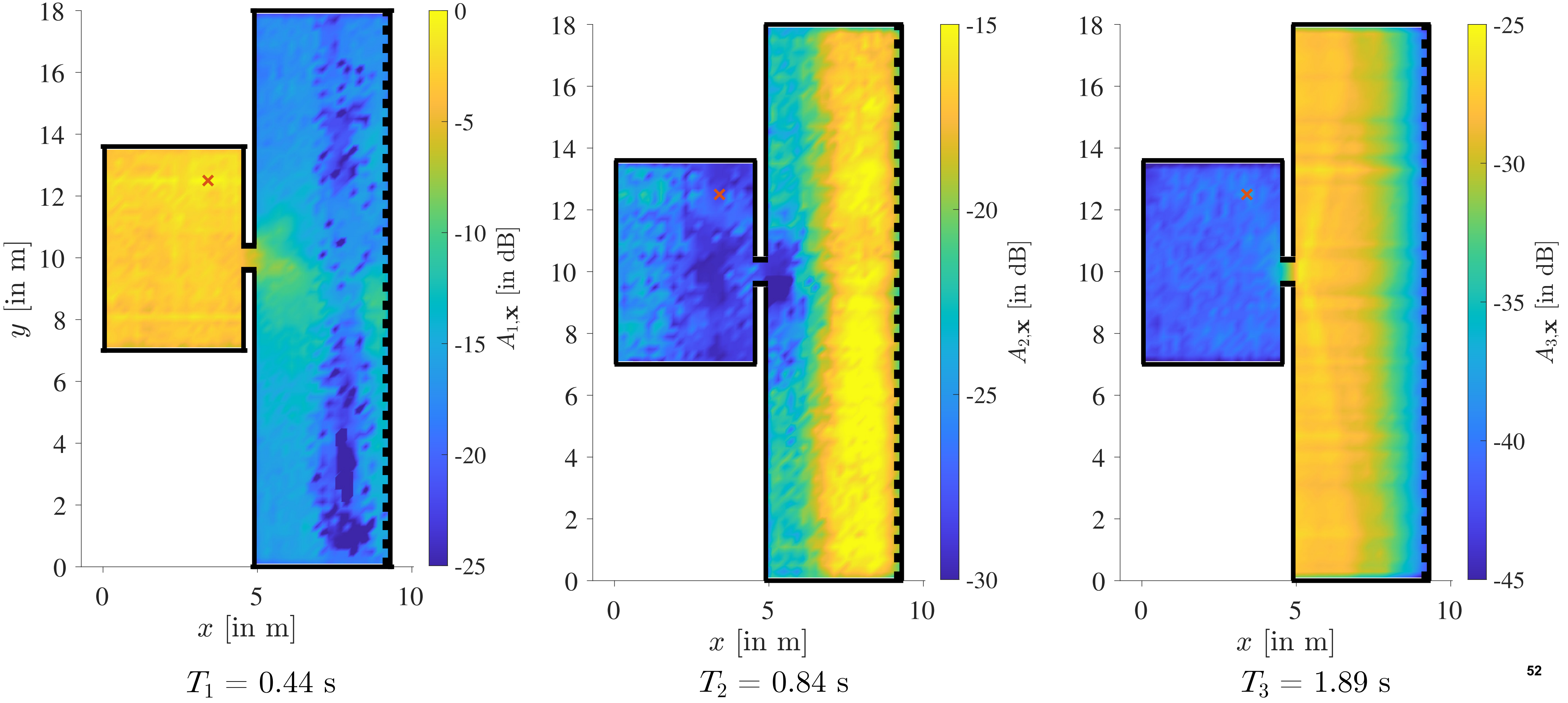


$T_1 = 0.46$ s

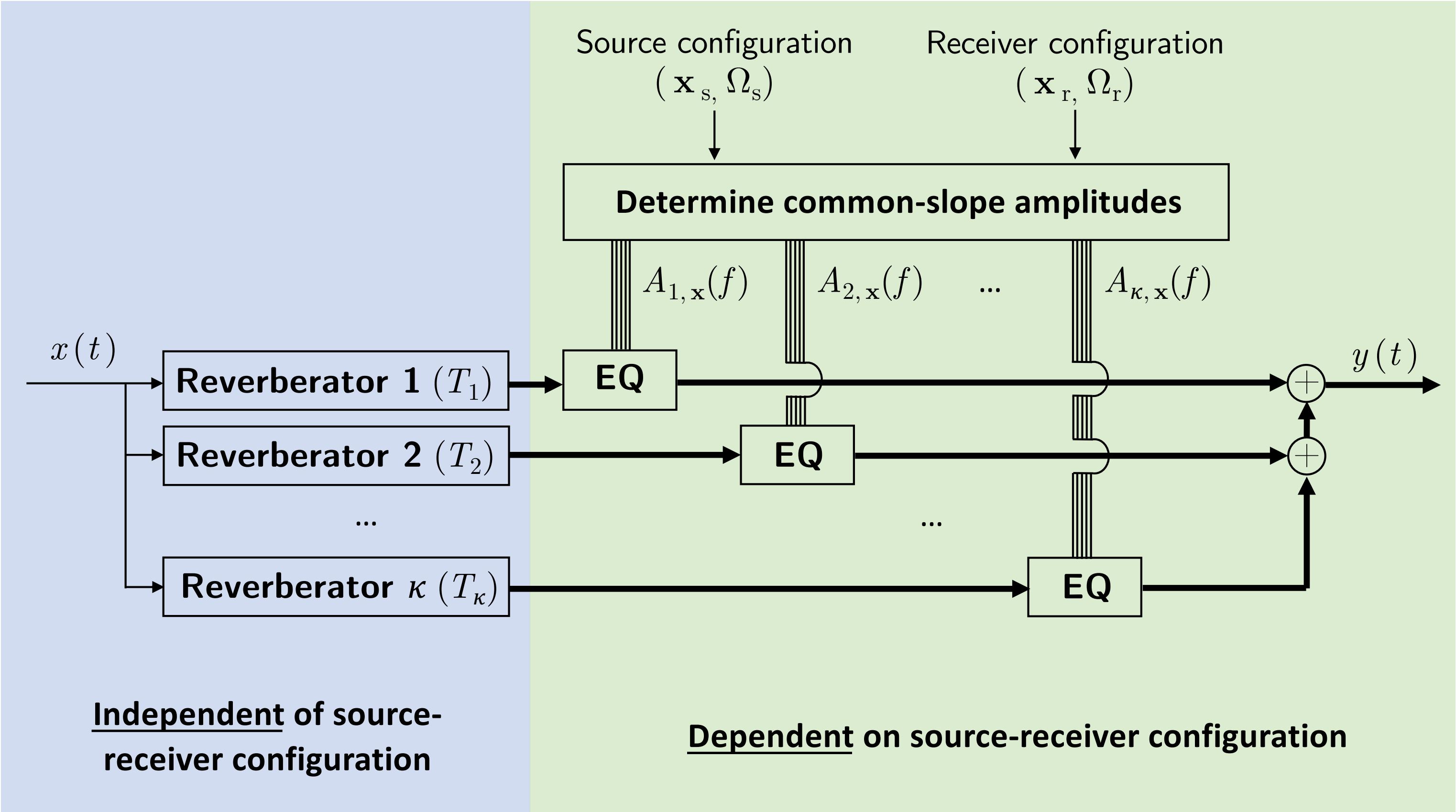
$T_2 = 1.51$ s

Mean dB-MSE = 0.36 dB

Common-slope analysis illustrates inhomogeneous sound energy decay due to non-uniform absorption



We can efficiently render inhomogeneous and anisotropic late reverberation using the common-slope model

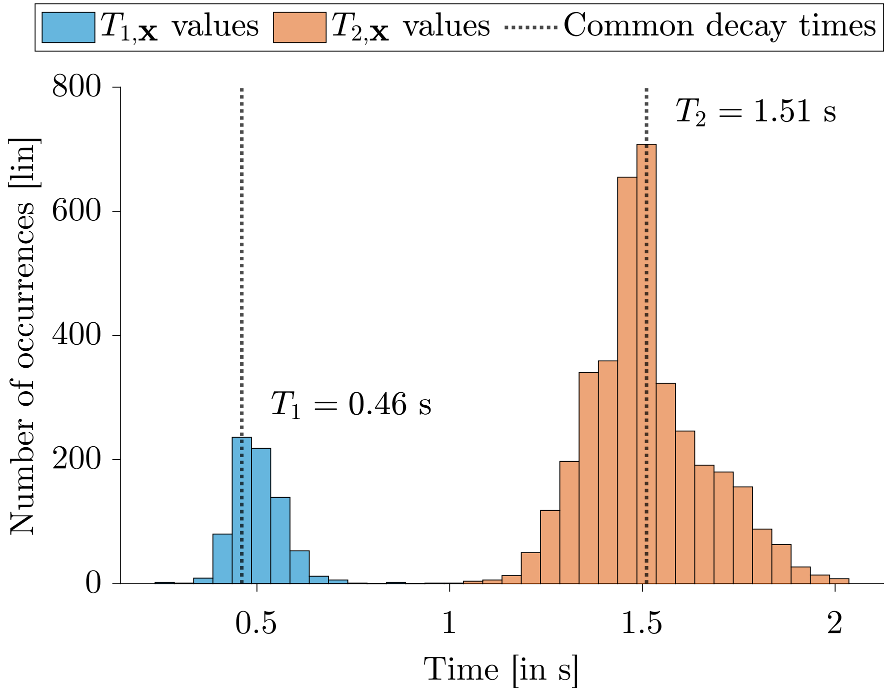


A!

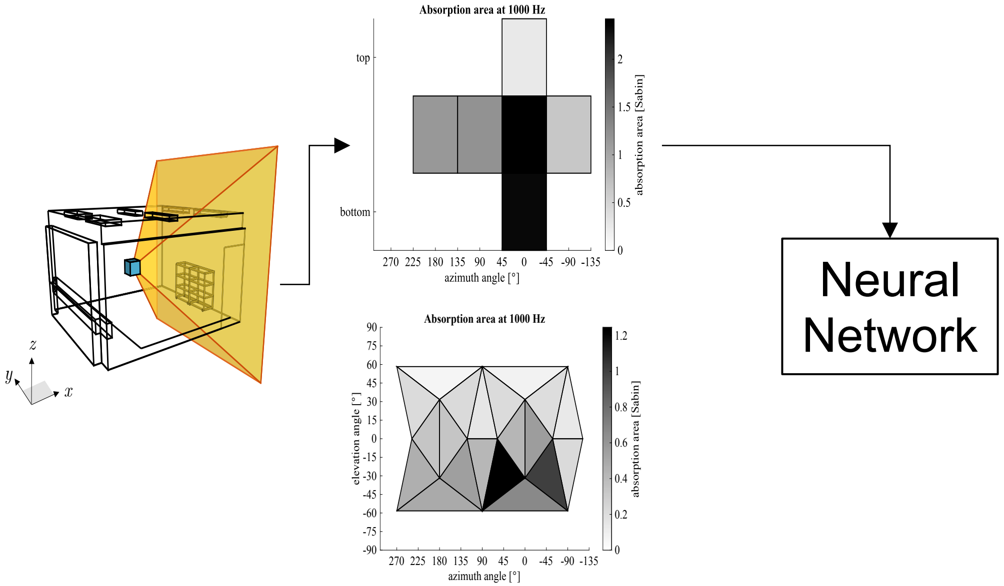
G. Götz, T. Kerimovs, S. J. Schlecht, and V. Pulkki. Dynamic late reverberation rendering using the common- slope model. Accepted for publication in *AES 6th International Conference on Audio for Games*, Tokyo, Japan, April 2024.

How to determine common-slopes and their amplitudes?

From a set of RIRs (measurements or simulations)

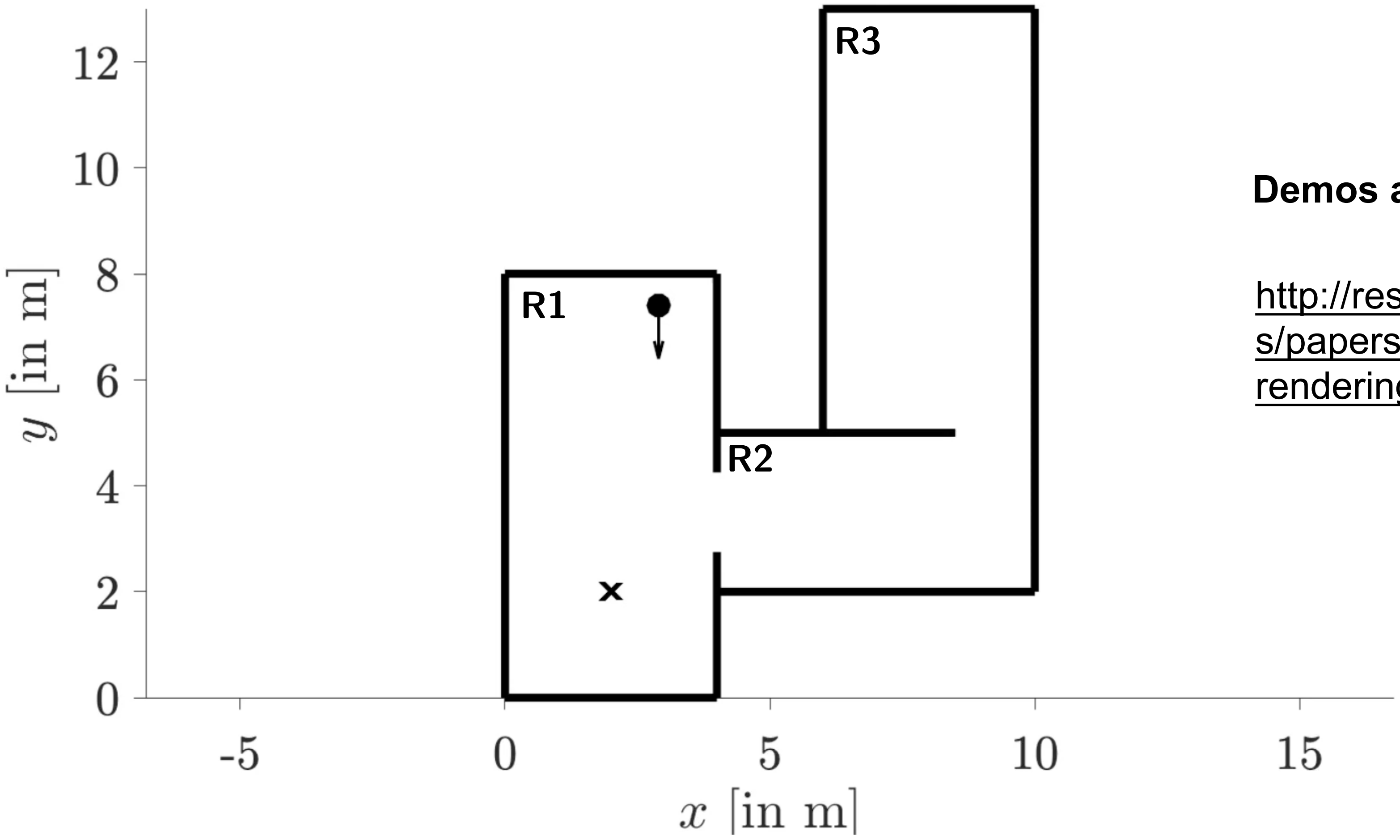


From a machine-learning-based approach (→ not done yet, future work)



Götz, G., Schlecht, S. J. & Pulkki, V. Common-slope modeling of late reverberation. *IEEEACM Trans. Audio, Speech, Lang. Process.* **31**, 3945–3957 (2023).

Demo: We can efficiently render inhomogeneous and anisotropic late reverberation using the common-slope model

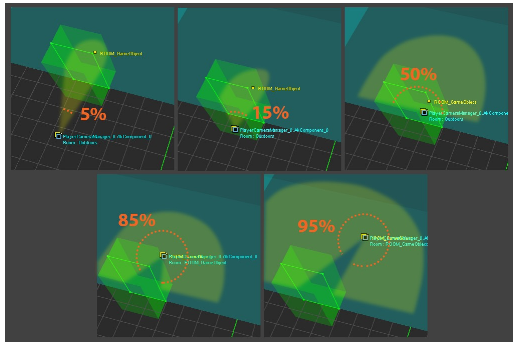


Demos available online:

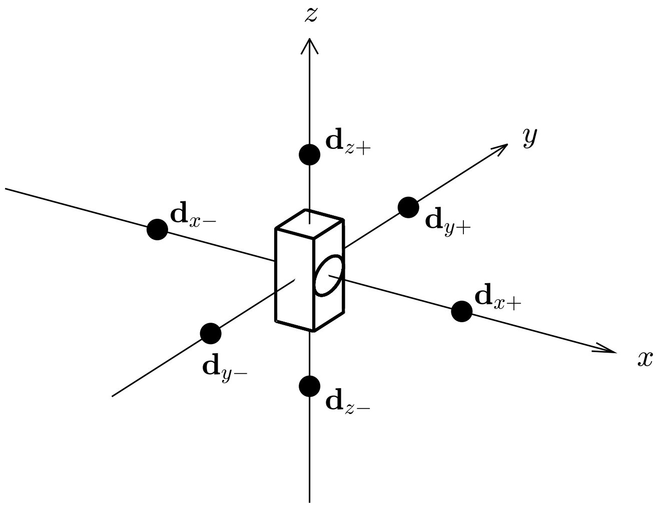
http://research.spa.aalto.fi/publications/papers/aes-games-common-slope-rendering/

A!

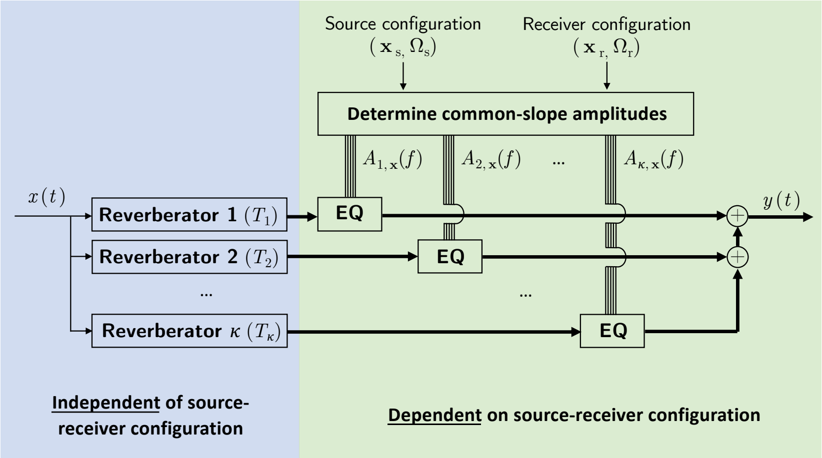
What you have learned today



Fundamentals of game audio engines



A simplified method for source directivity rendering



An efficient and data-driven approach for late reverberation rendering

A!

Kiitos
aalto.fi