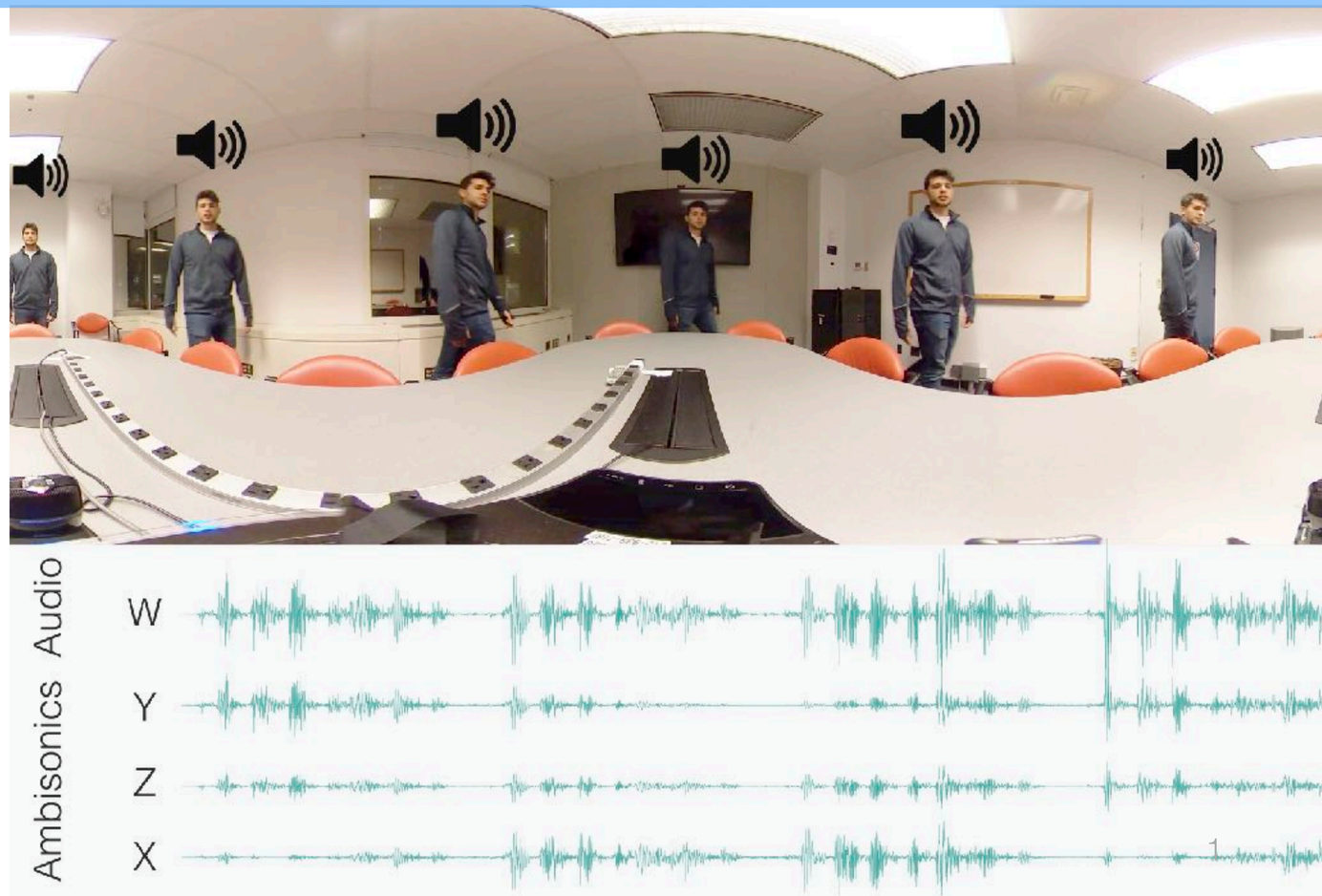
Scene-Aware Audio for 360° Videos

Dingzeyu Li
Timothy R. Langlois
Changxi Zheng

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Adobe Research



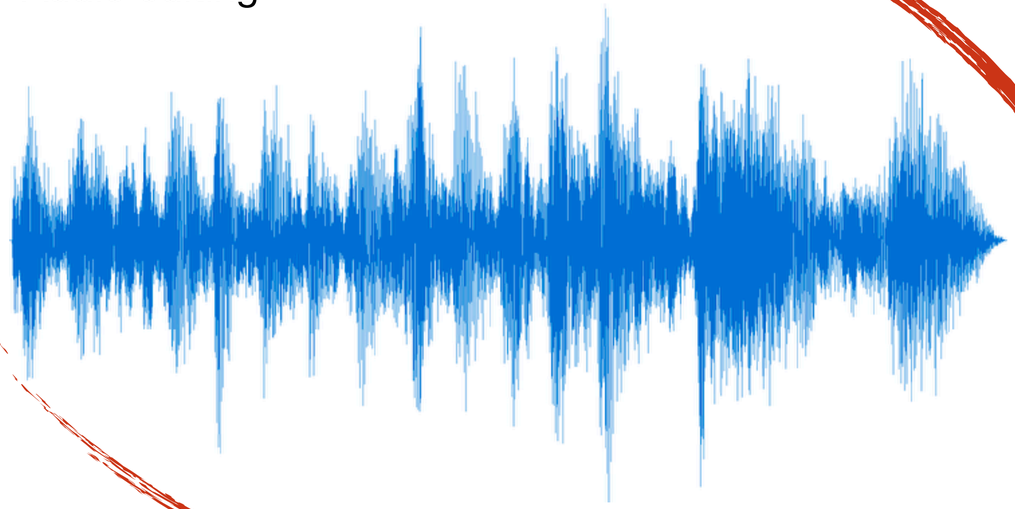
360° Video Ecosystem

360° cameras

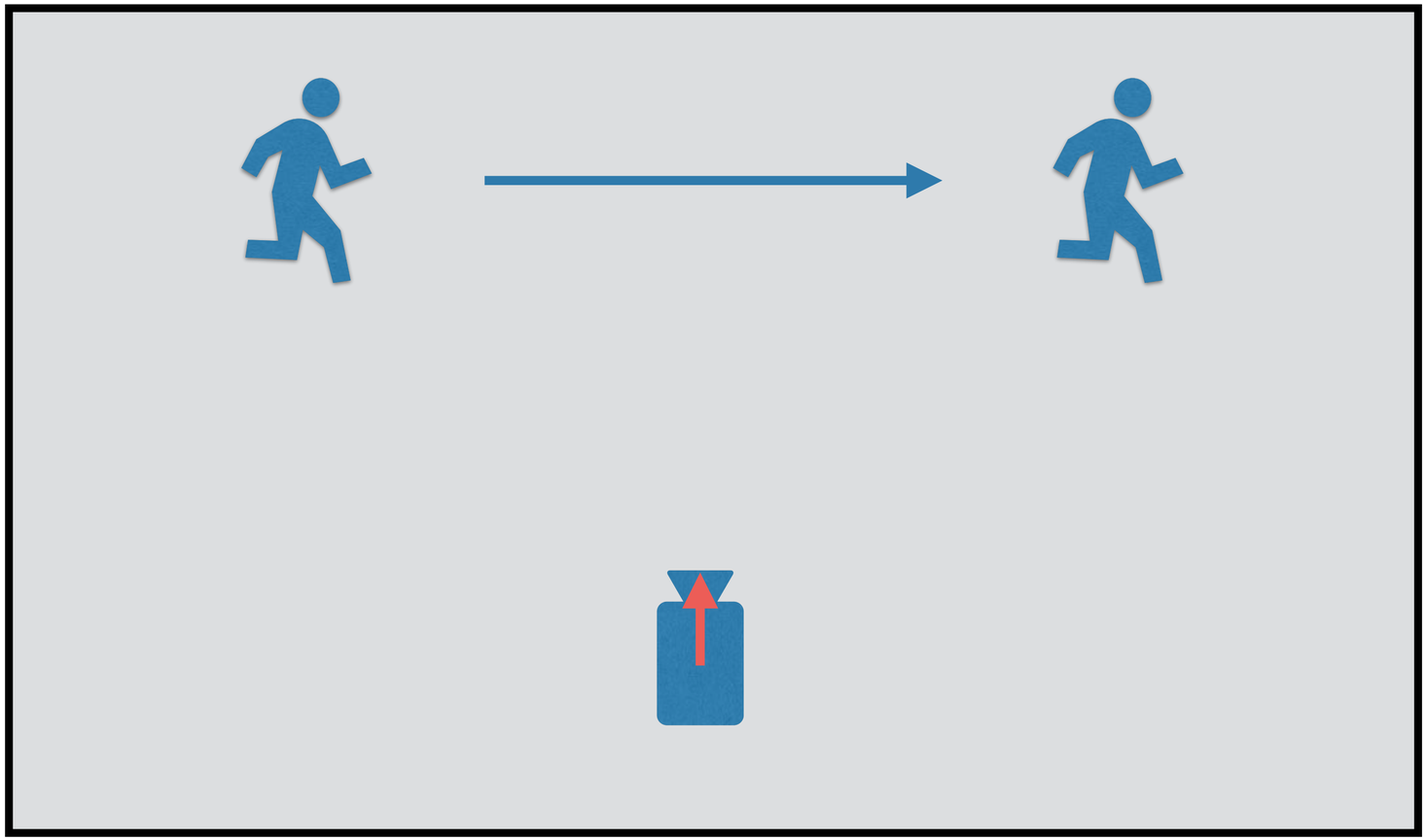
Playback devices

Video editing

Audio editing

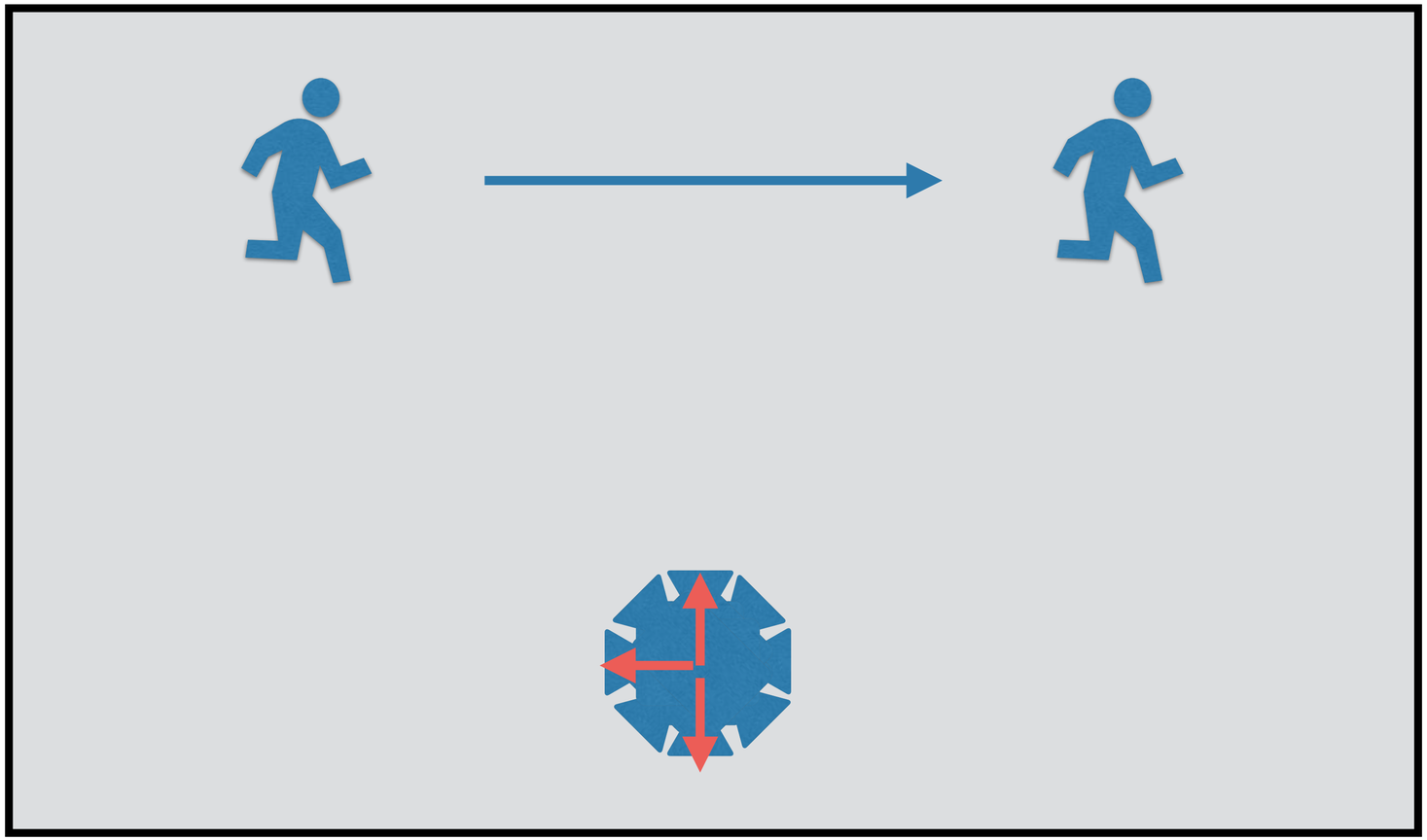


Challenge: Traditional Stereo Audio



fixed viewing direction

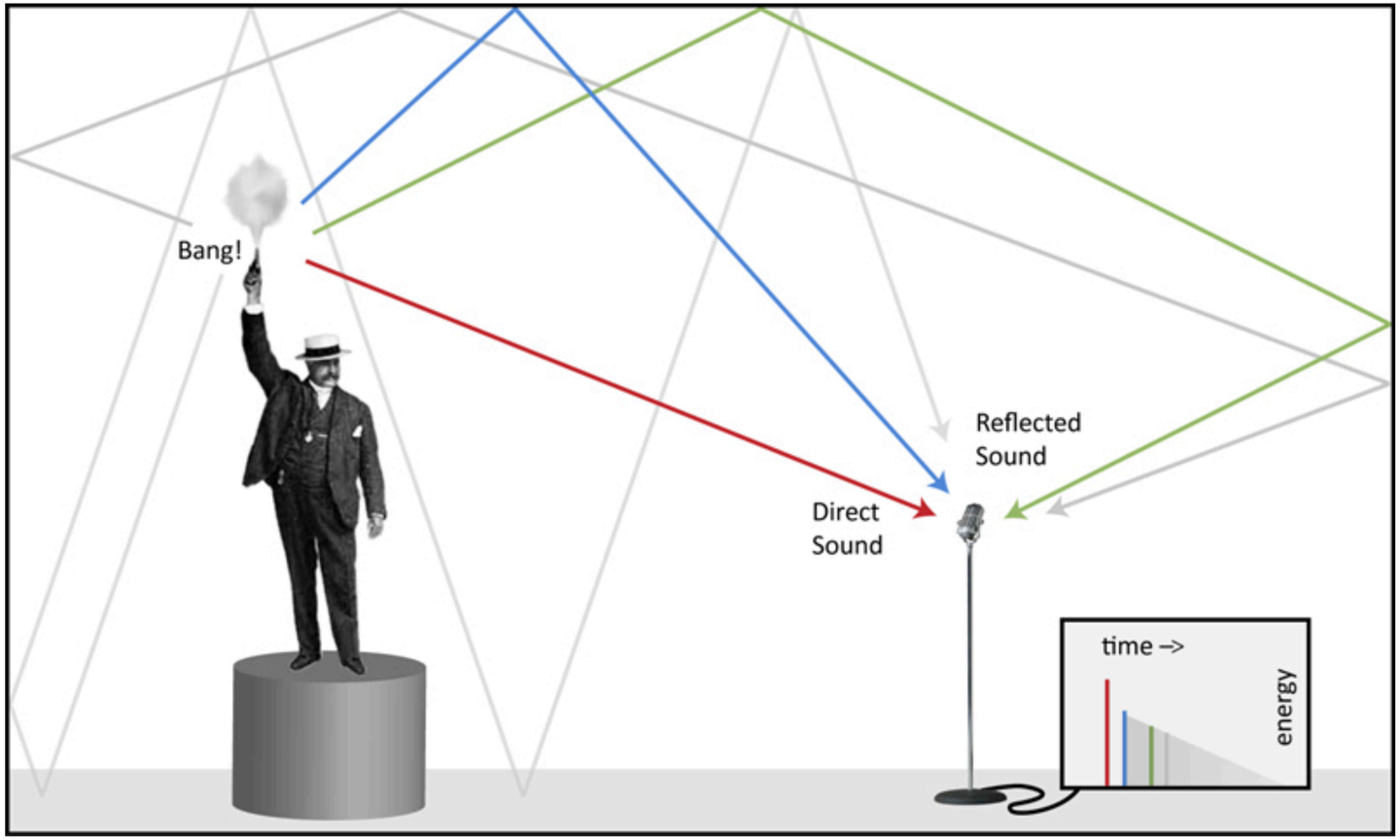
Challenge: 360° Spatial Audio



varying viewing direction

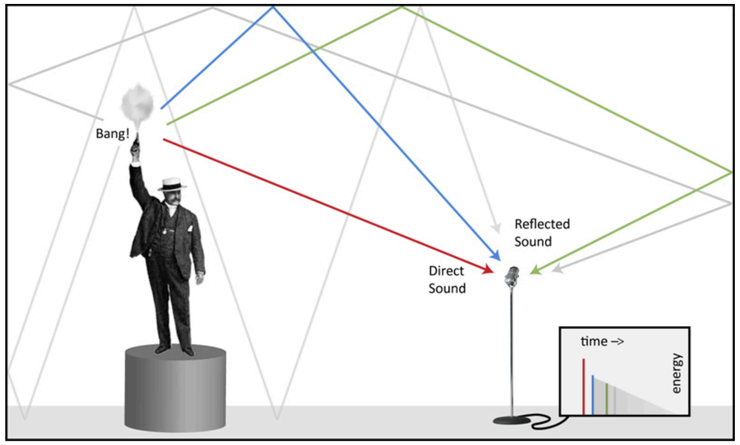
Room Acoustics

Impulse Response (IR)

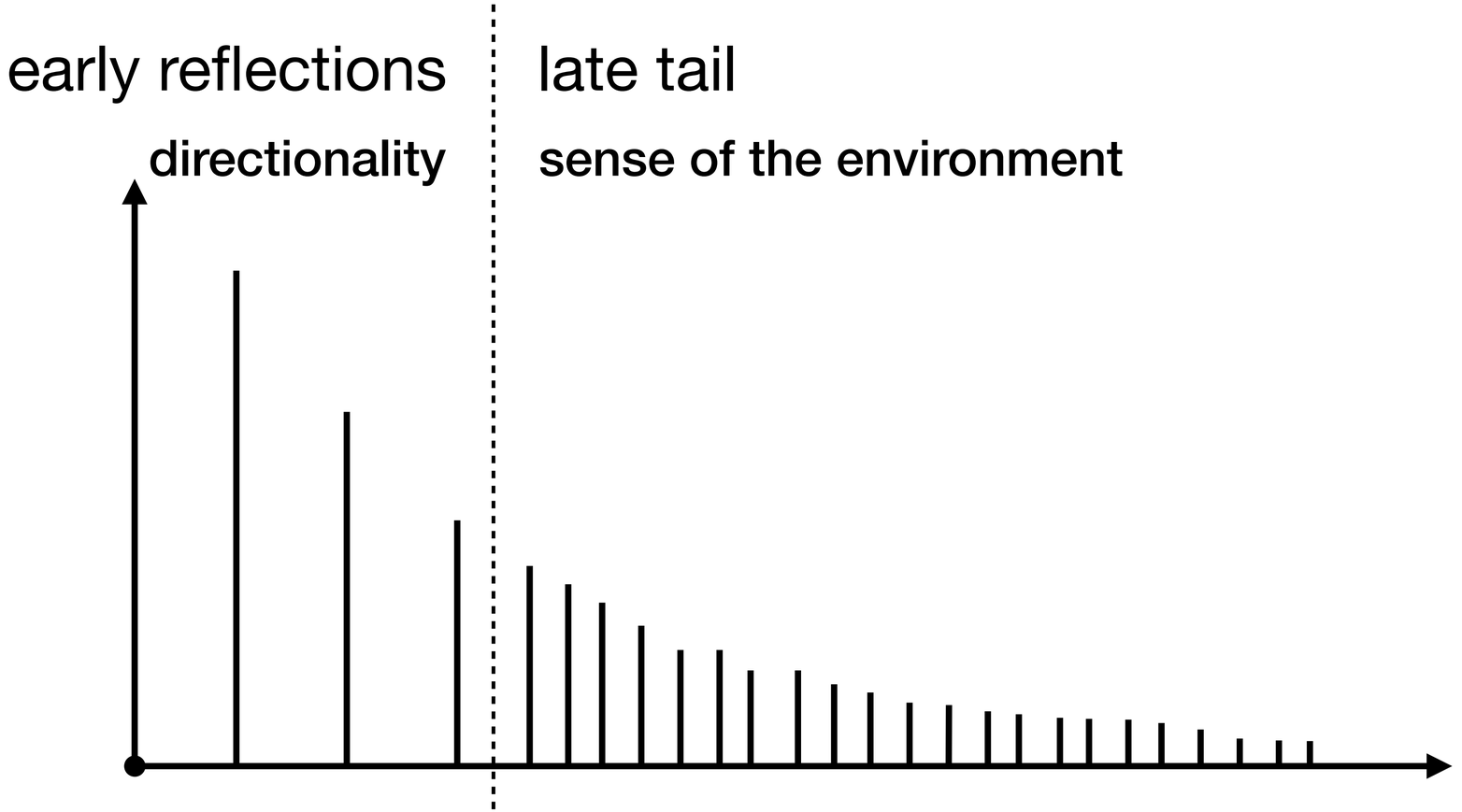


source: rational acoustics

Impulse Response (IR)



source: rational acoustics

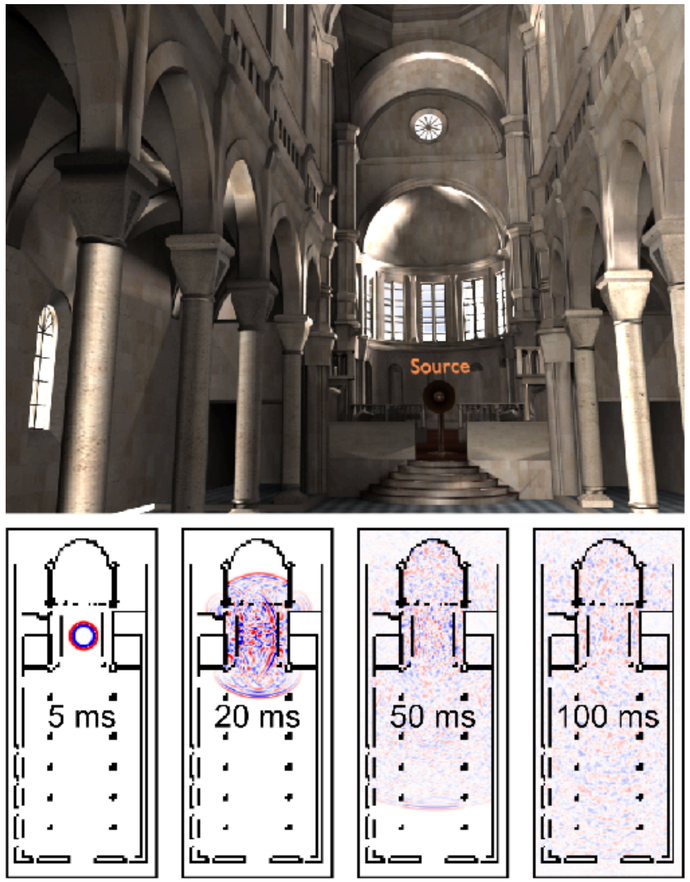


IR from recording

- easy to setup and obtain
- lack the dynamic directional effects

[Zotkin et al. 2004]
[Tervo et al. 2010]
[Traer and McDermott 2016]

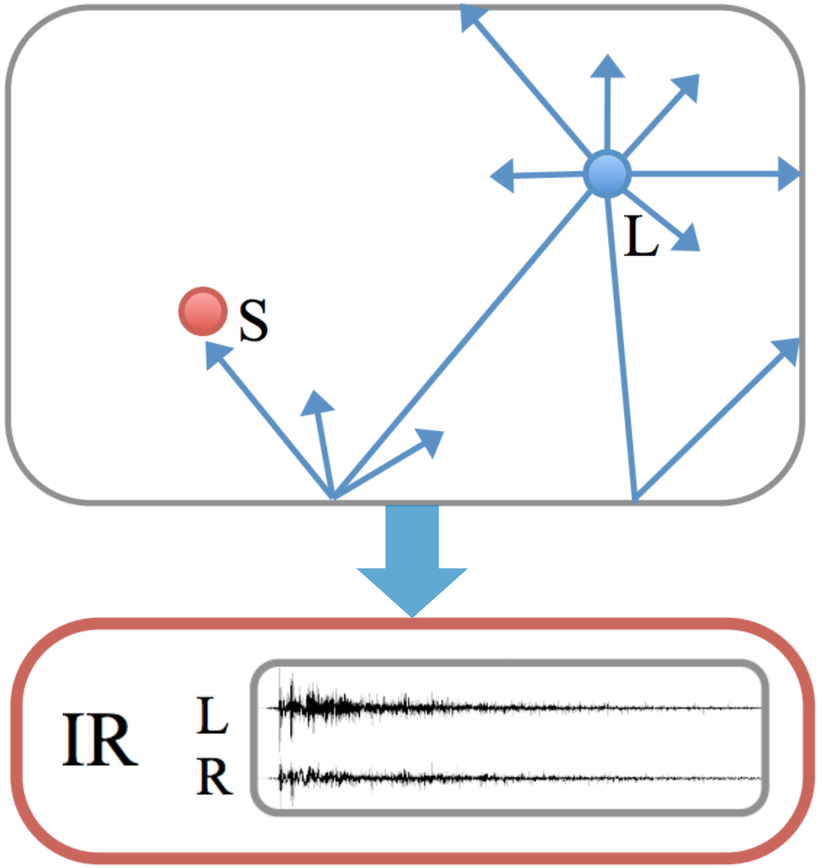
IR from Acoustic Simulation



[Raghuvanshi et al. 2009]

wave-based: accurate

slow to simulate

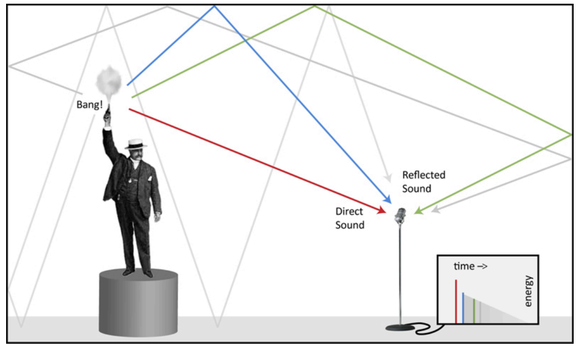


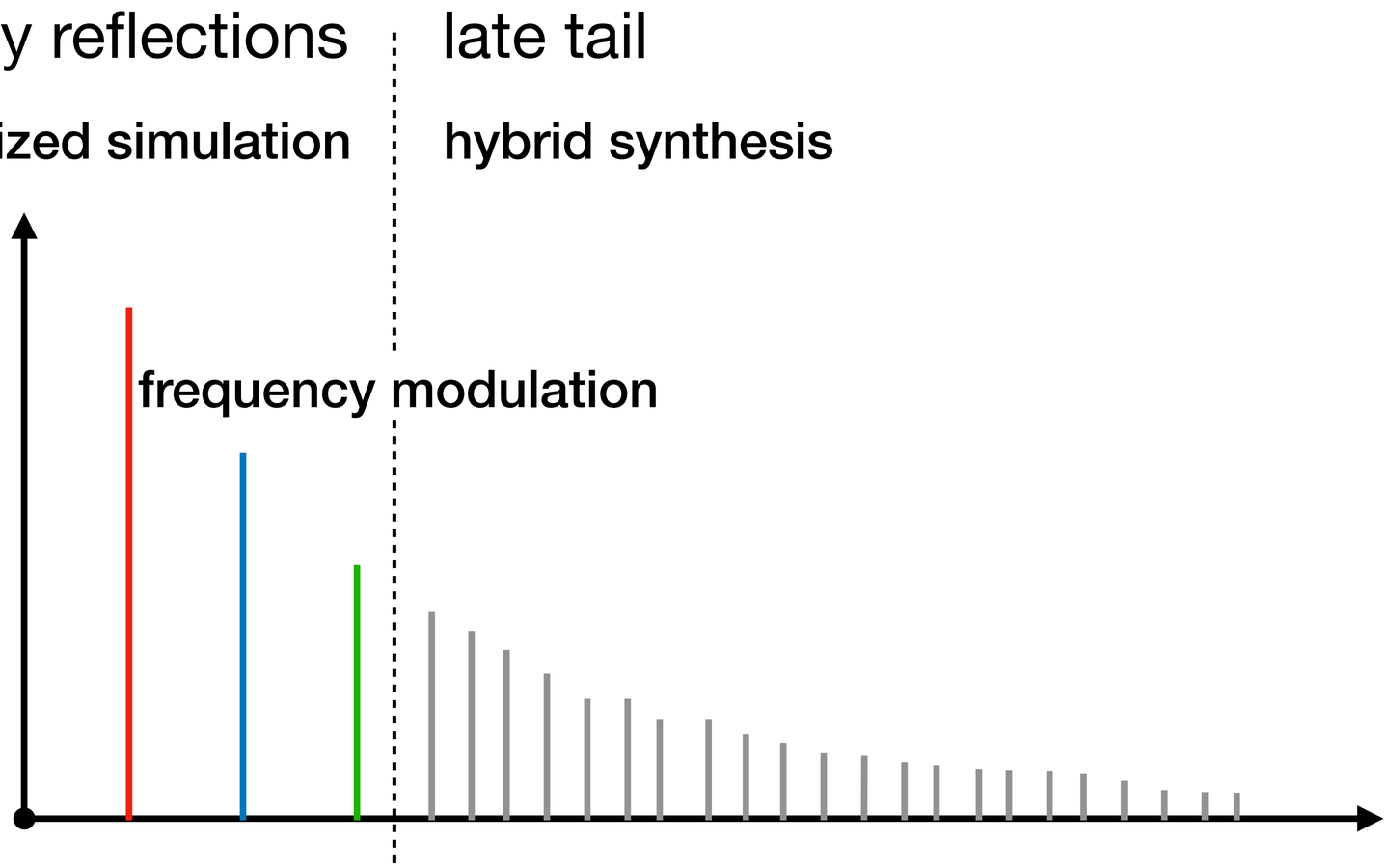
[Schissler et al. 2017]

geometric acoustics: fast

lack wave-based behavior

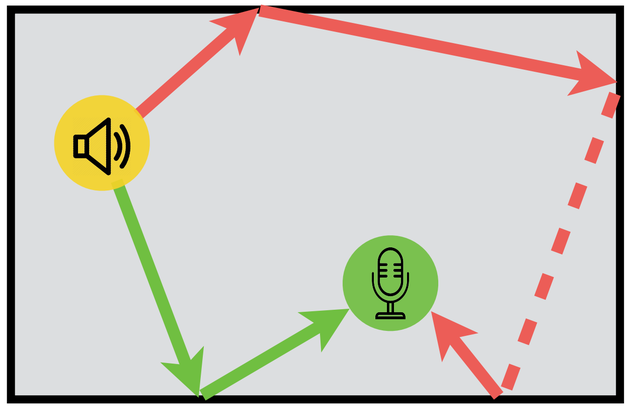
Key Insight: Isotropy in late IR



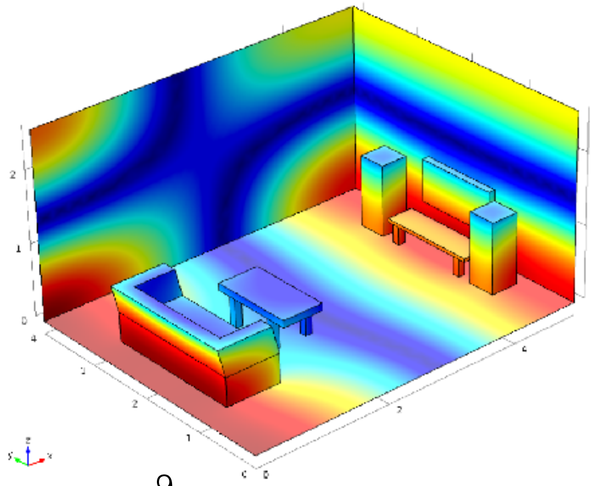


Contributions

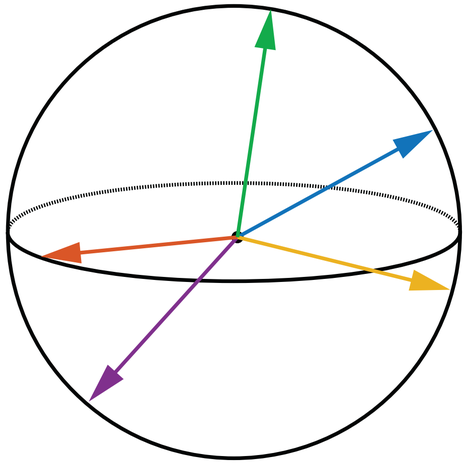
room material optimization



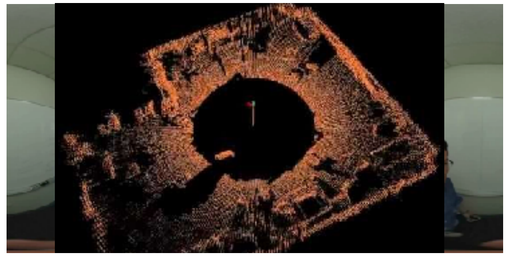
frequency modulation

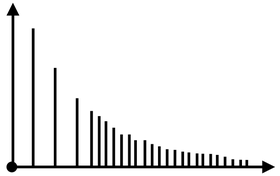


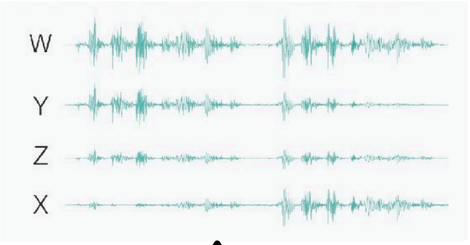
hybrid synthesis



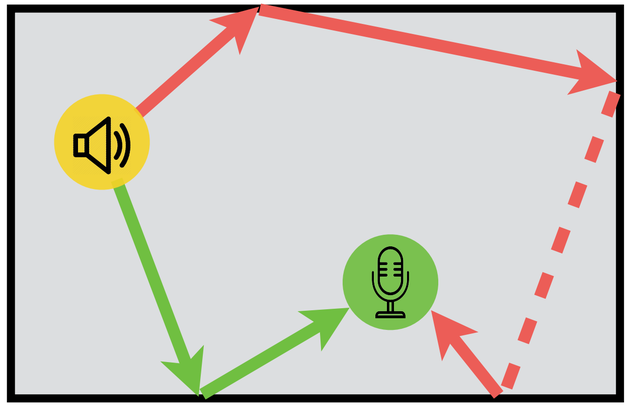
Scene-Aware Spatial Audio Pipeline

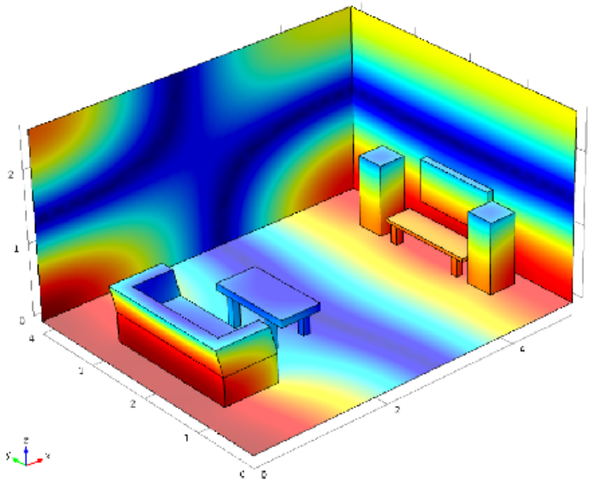


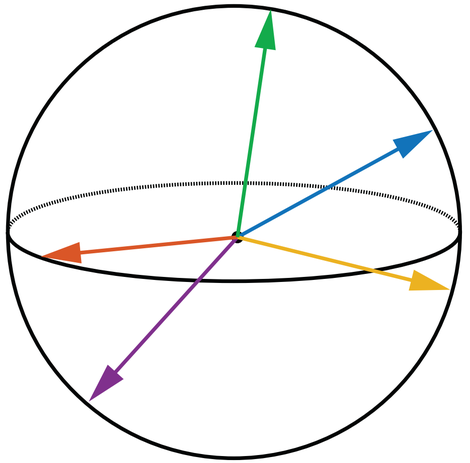




room material optimization → frequency modulation → hybrid synthesis



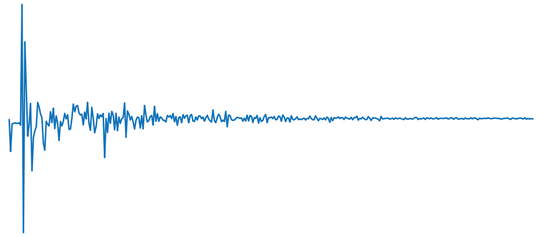


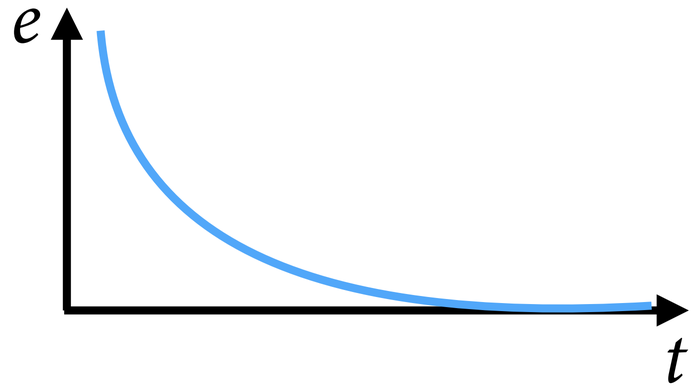


Acoustic Measurement

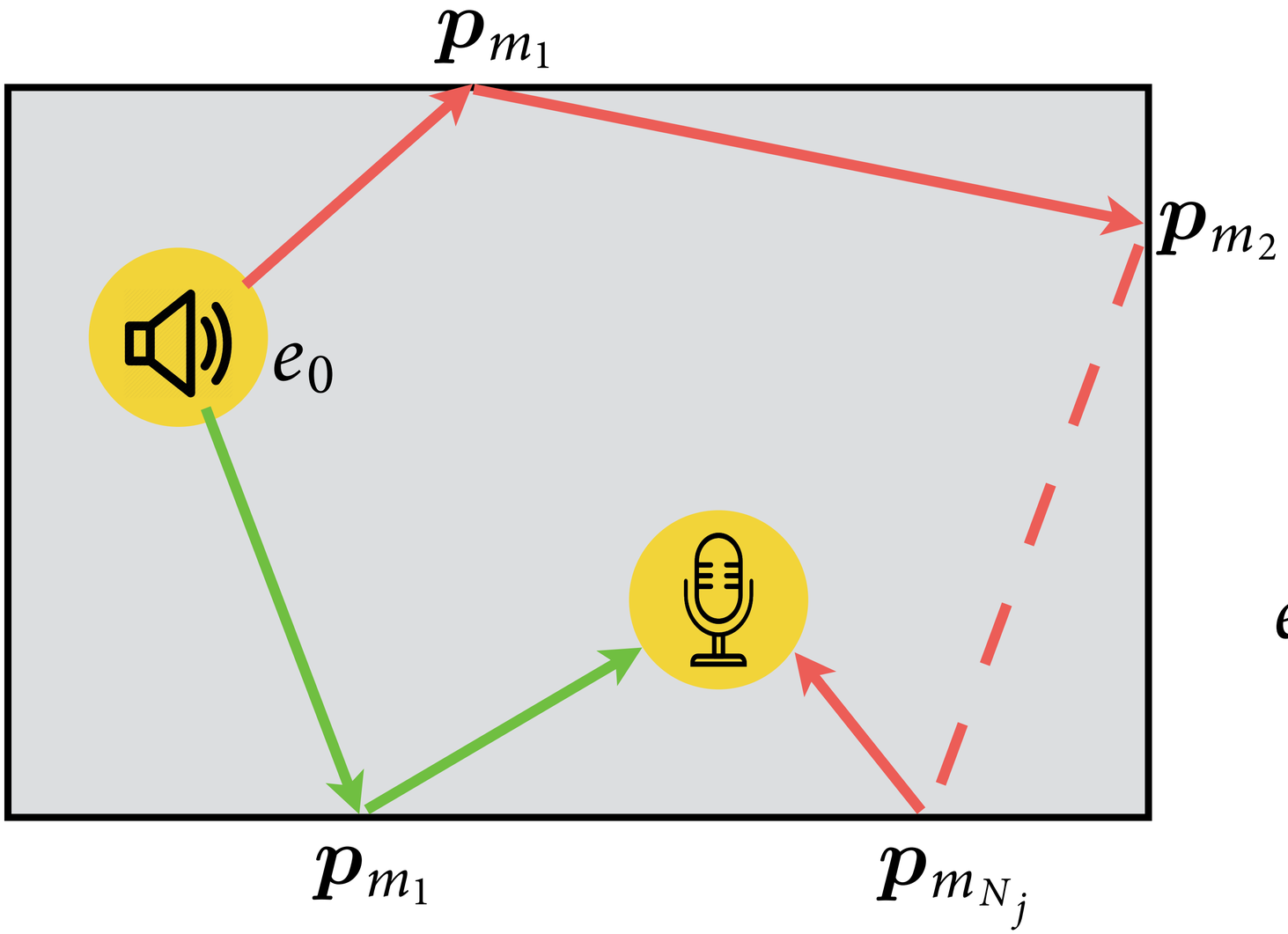
Measurement Setup

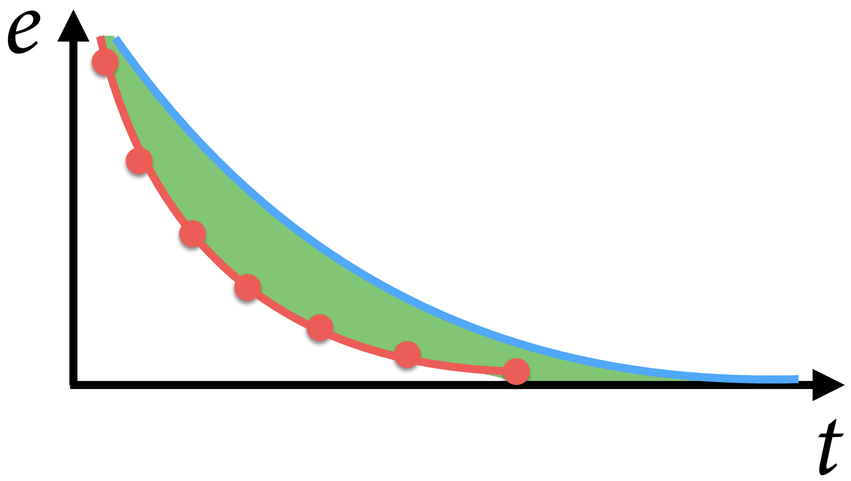
Impulse Response



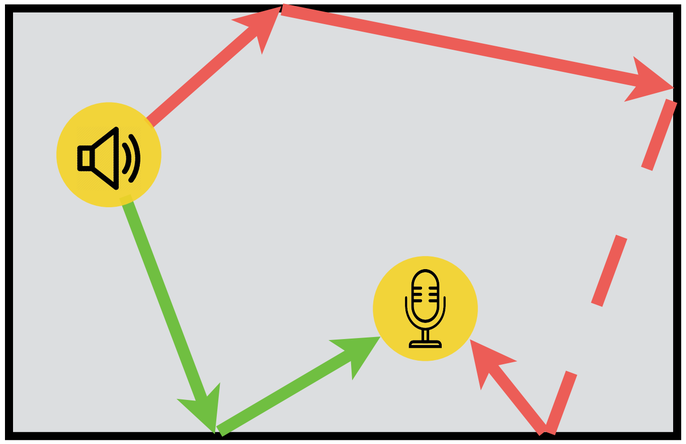


Geometric Acoustic Simulation





Material Optimization



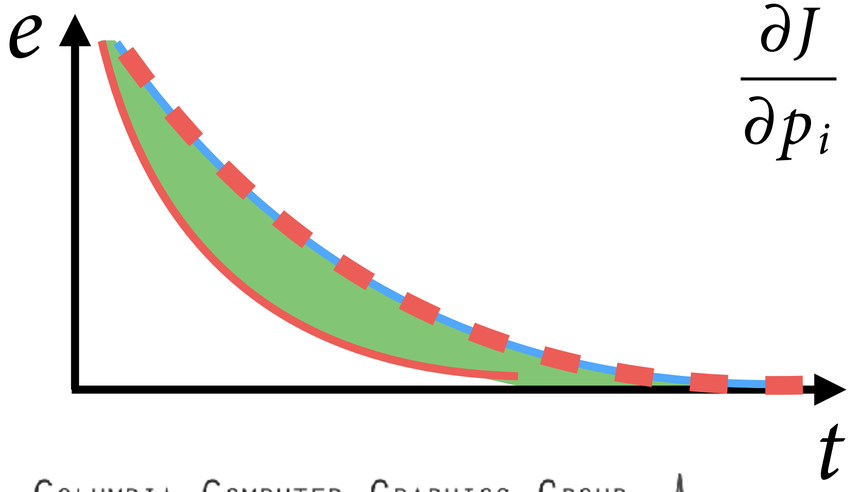
$$e_j(\boldsymbol{p}) = e_0 \prod_{i=1}^{N_j} \boldsymbol{p}_{m(i)}$$

Objective Function

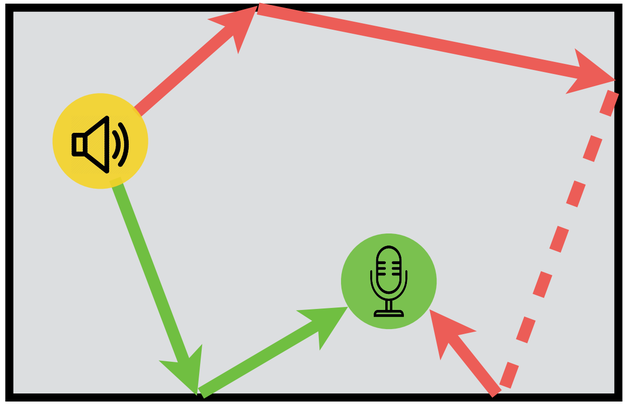
$$J(\boldsymbol{p}) = \sum_{j=1}^{M} \left[\log_{10}\left(\frac{e_j(\boldsymbol{p})}{e_0}\right) - \log_{10}\left(\frac{\tilde{h}(t_j)}{\tilde{h}(\bar{t}_0)}\right) \right]^2$$

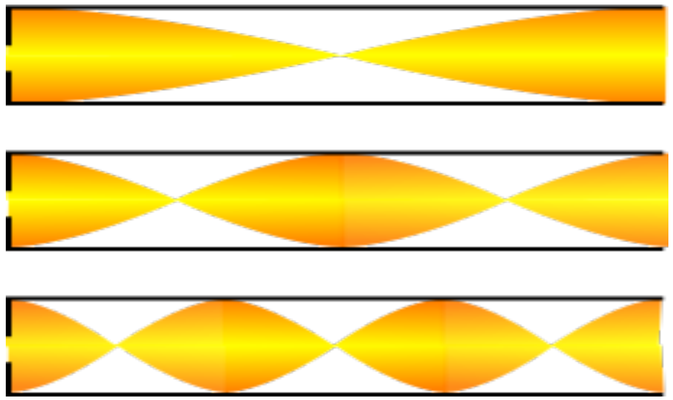
$$\frac{\partial J}{\partial p_i} = \frac{2}{\ln 10} \sum_{j=1}^{M} \left[\log_{10}\left(\frac{e_j}{e_0}\right) - \log_{10}\left(\frac{\tilde{h}(t_j)}{\tilde{h}(\bar{t}_0)}\right) \right] \frac{\partial e_j}{\partial p_i}$$

- analytic gradient
- 20x speedup



Limitation of Geometric Acoustics





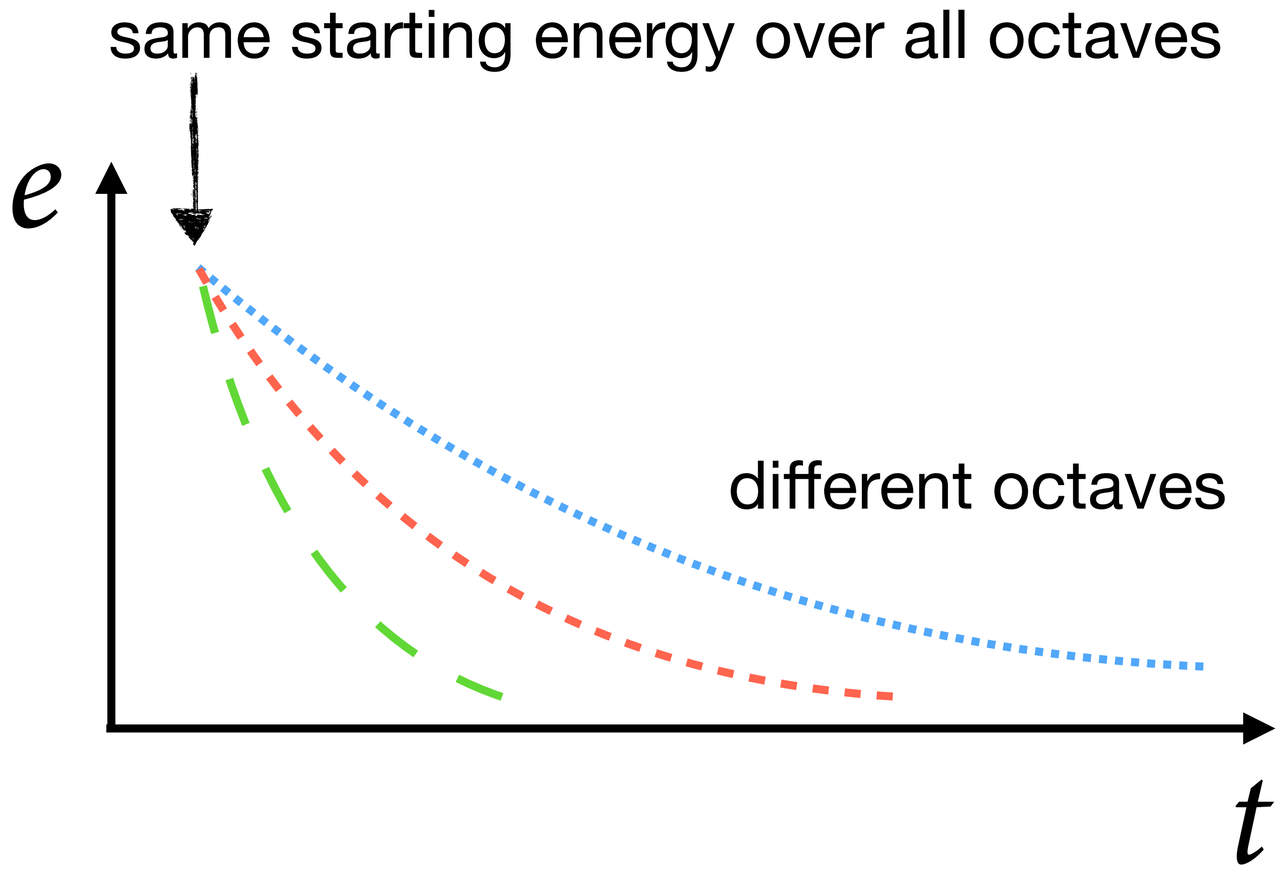
light rays ~ sound waves

inaccurate approximation for low frequencies

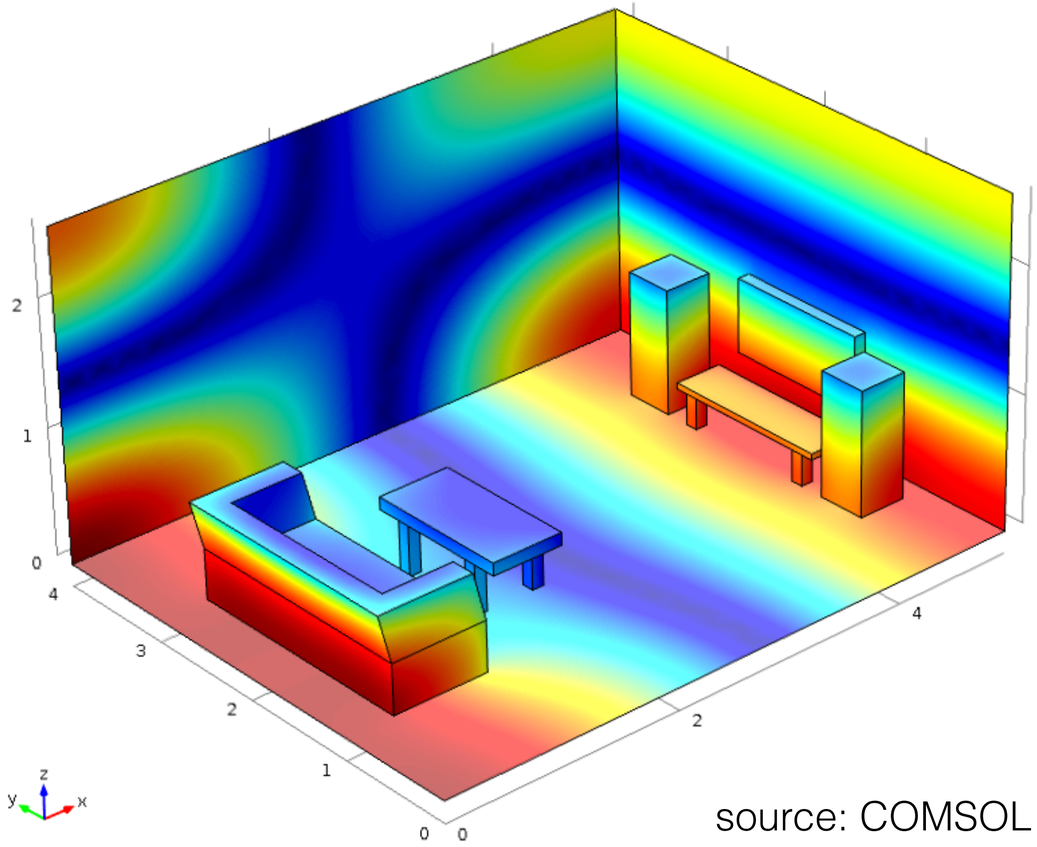
rooms are acoustic filters.

resonant frequencies and modes

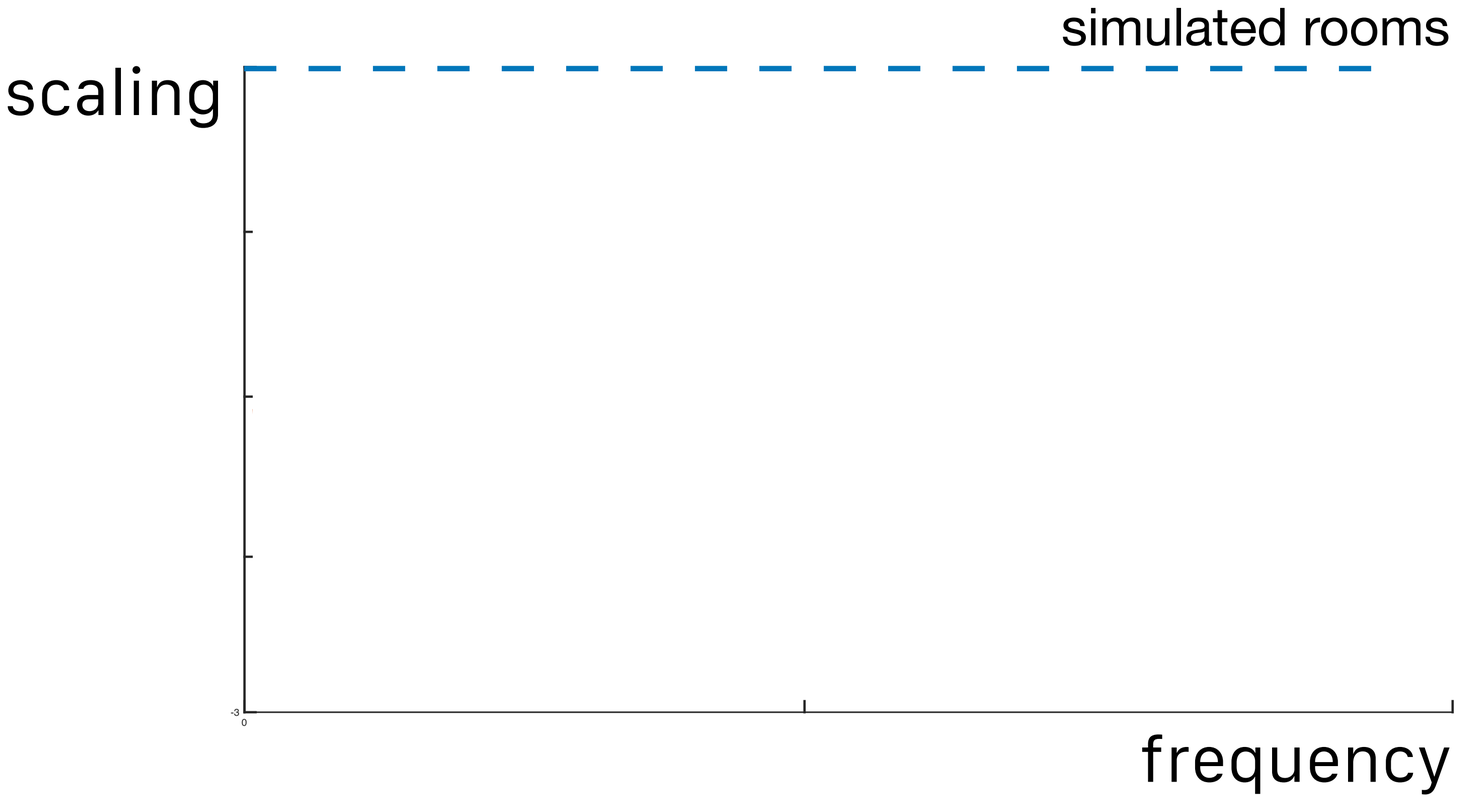
Geometric Acoustics for Different Frequencies



Real-World Frequency Modulation



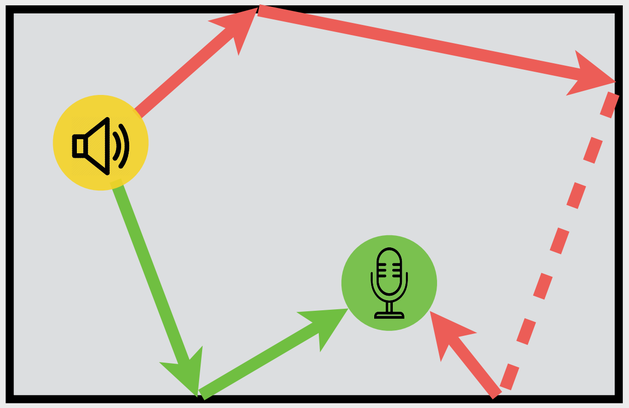
source: COMSOL

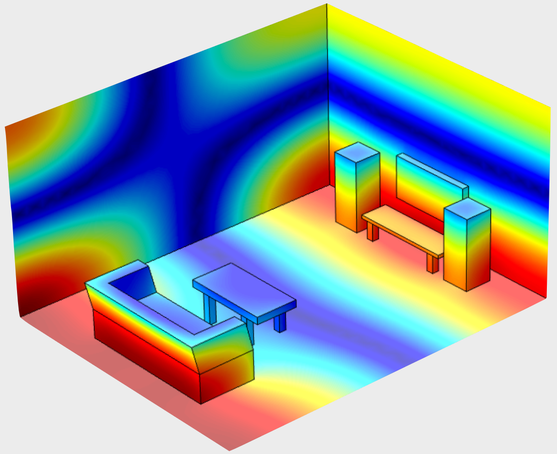


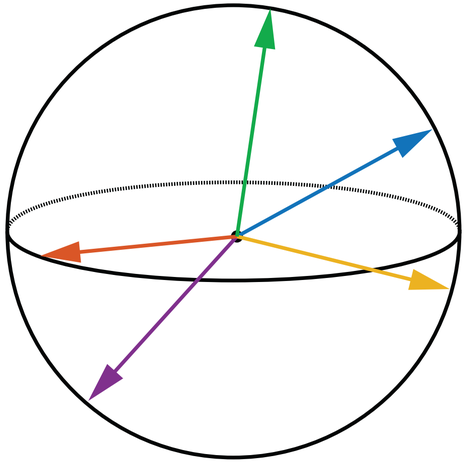
equalization matching [Germain et al. 2016]

Pipeline Recap

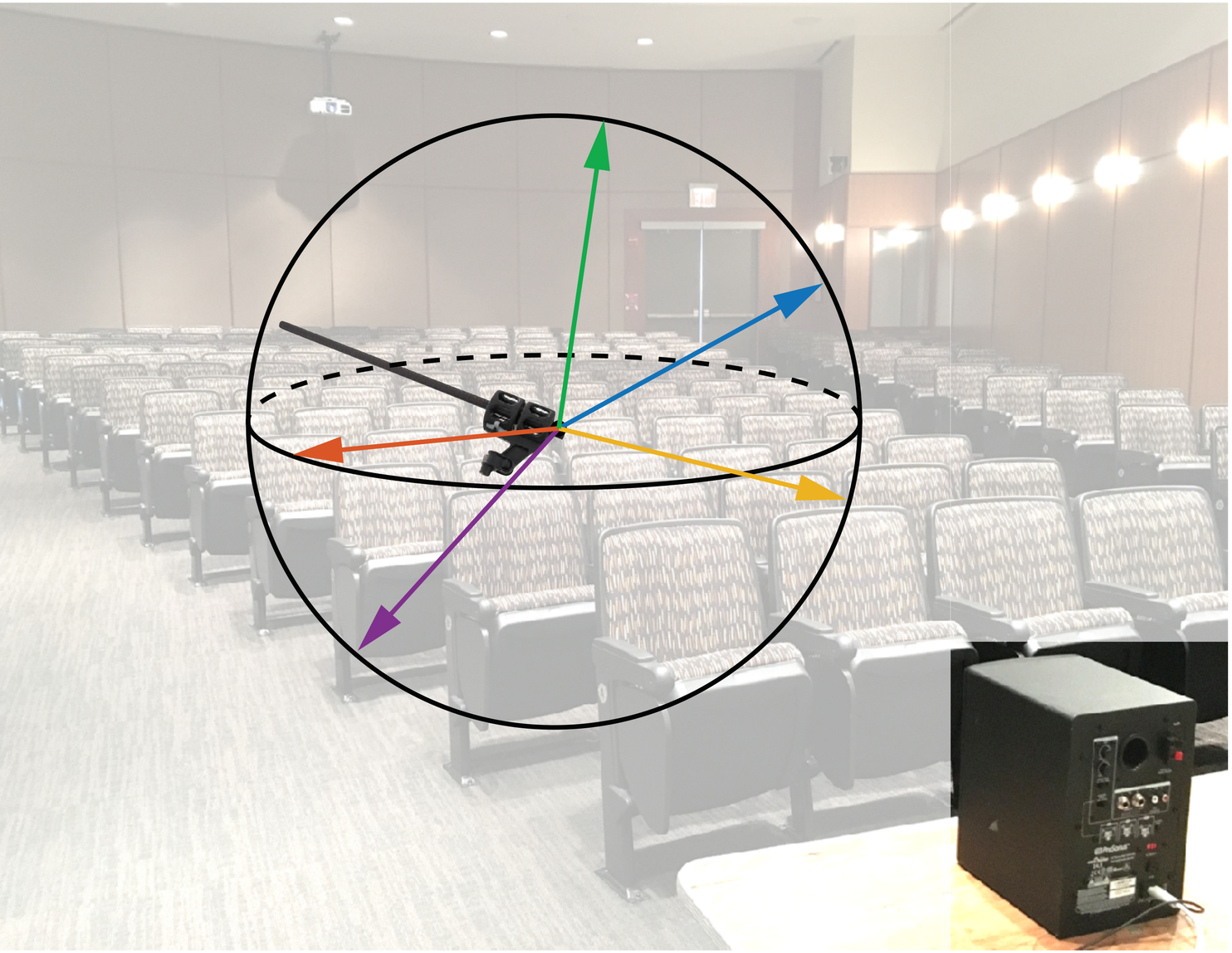
room material optimization → frequency modulation → hybrid synthesis

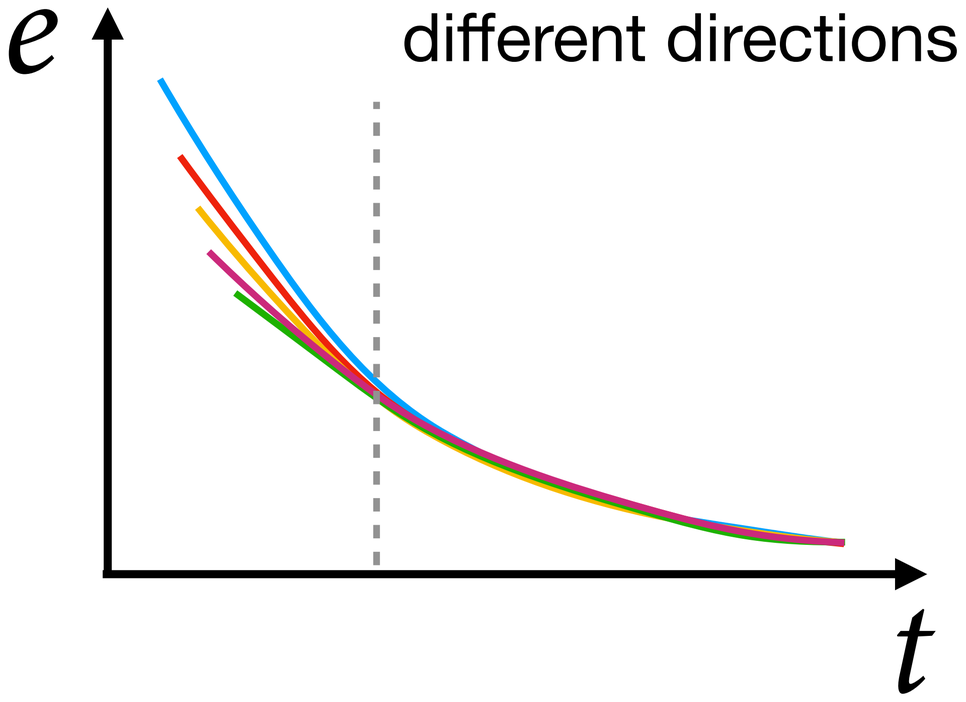






Isotropy Measurement

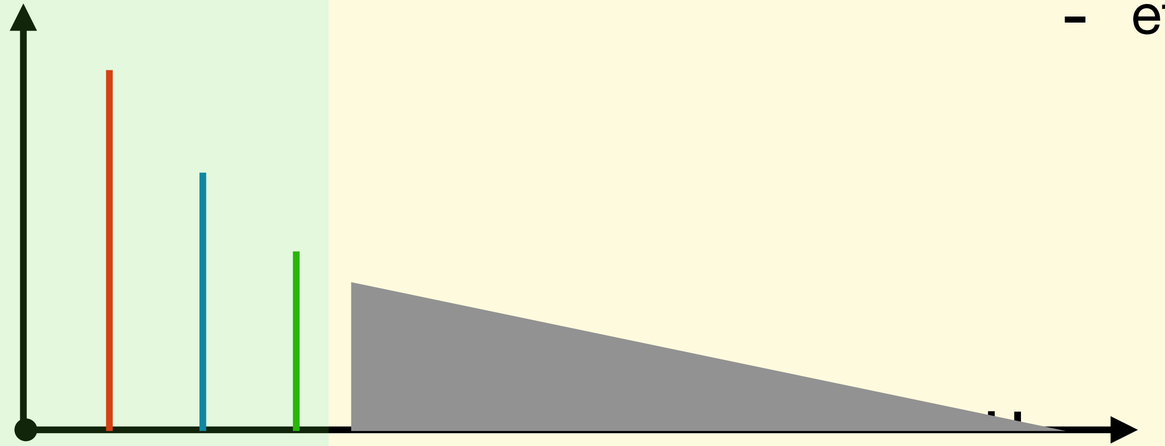




Hybrid Synthesis

Optimized Early Reflections

- fast simulation
- directionality



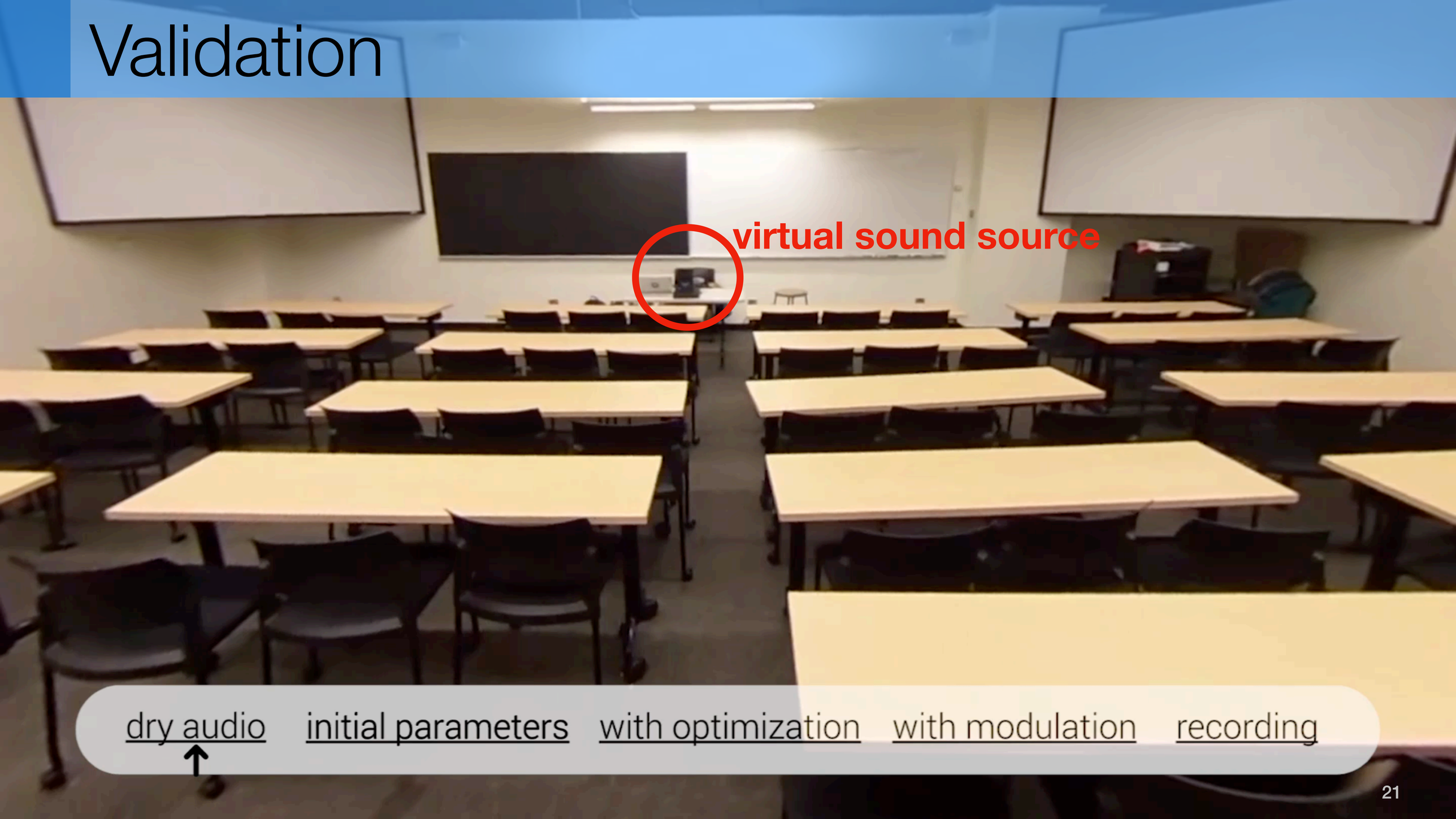
Late Tail

- summation over diffuse tail
- easy implementation
- efficient computation

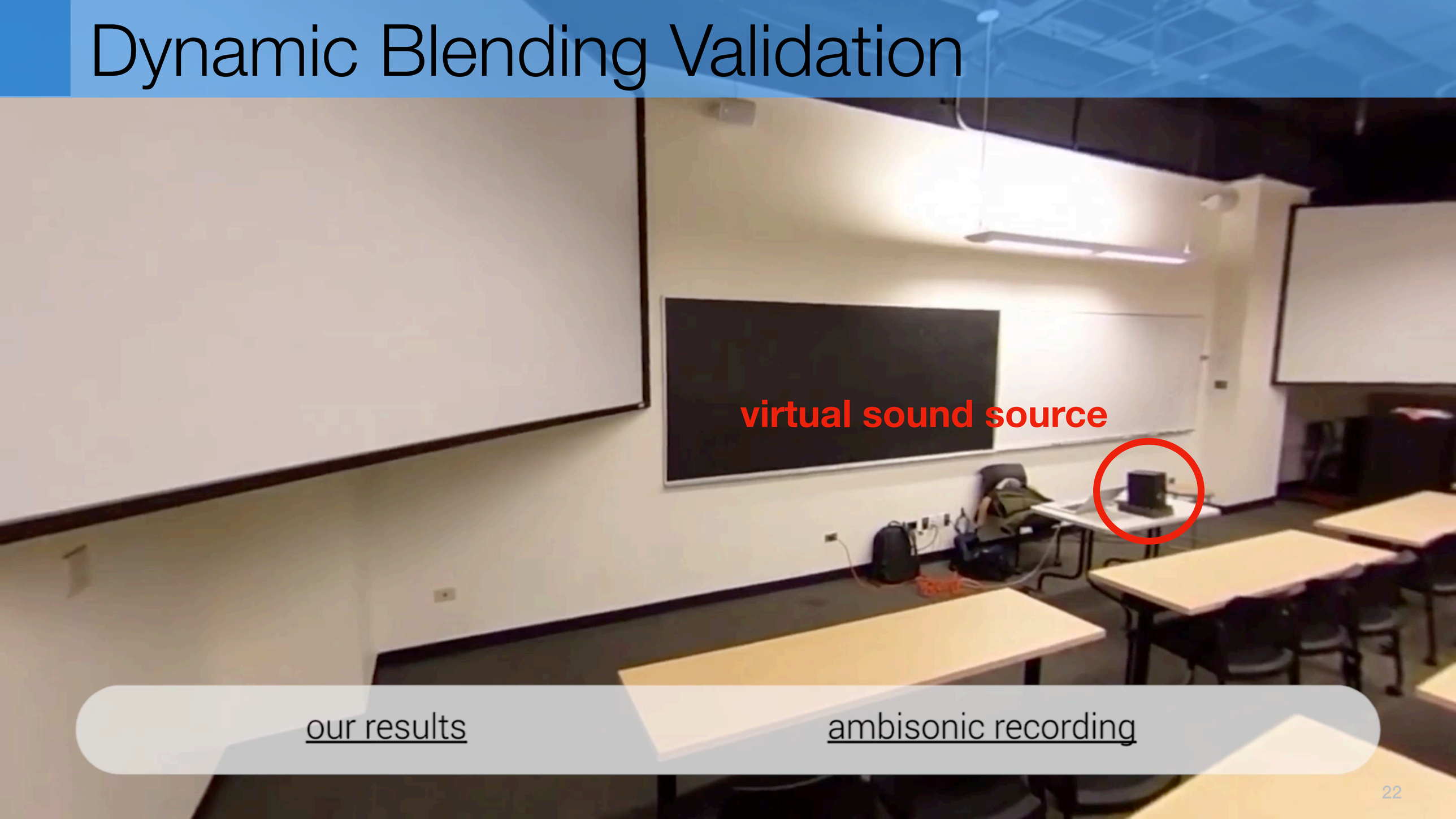
Validation and Applications

headphones recommended

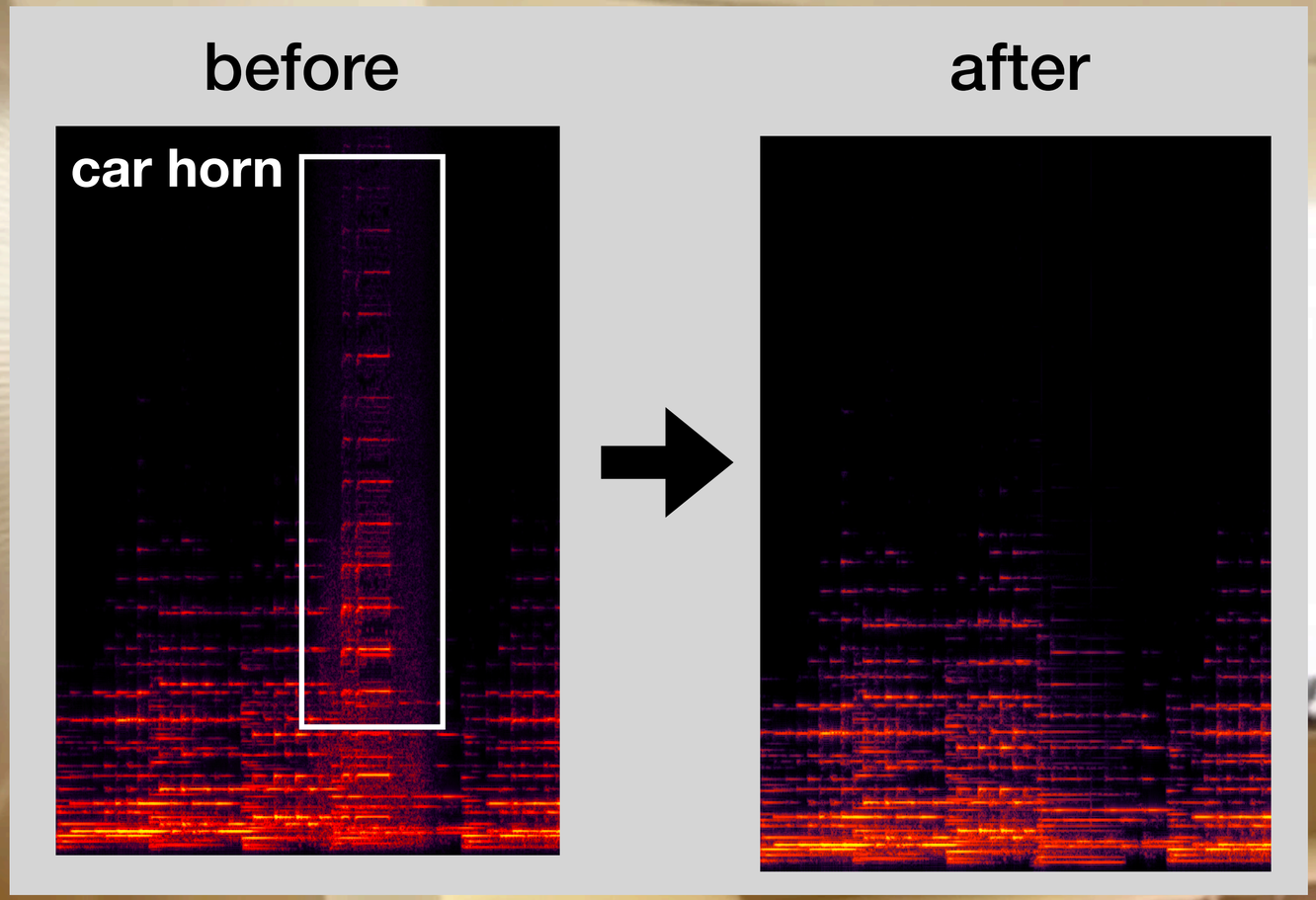
Validation


virtual sound source
dry audio
initial parameters
with optimization
with modulation
recording

Dynamic Blending Validation



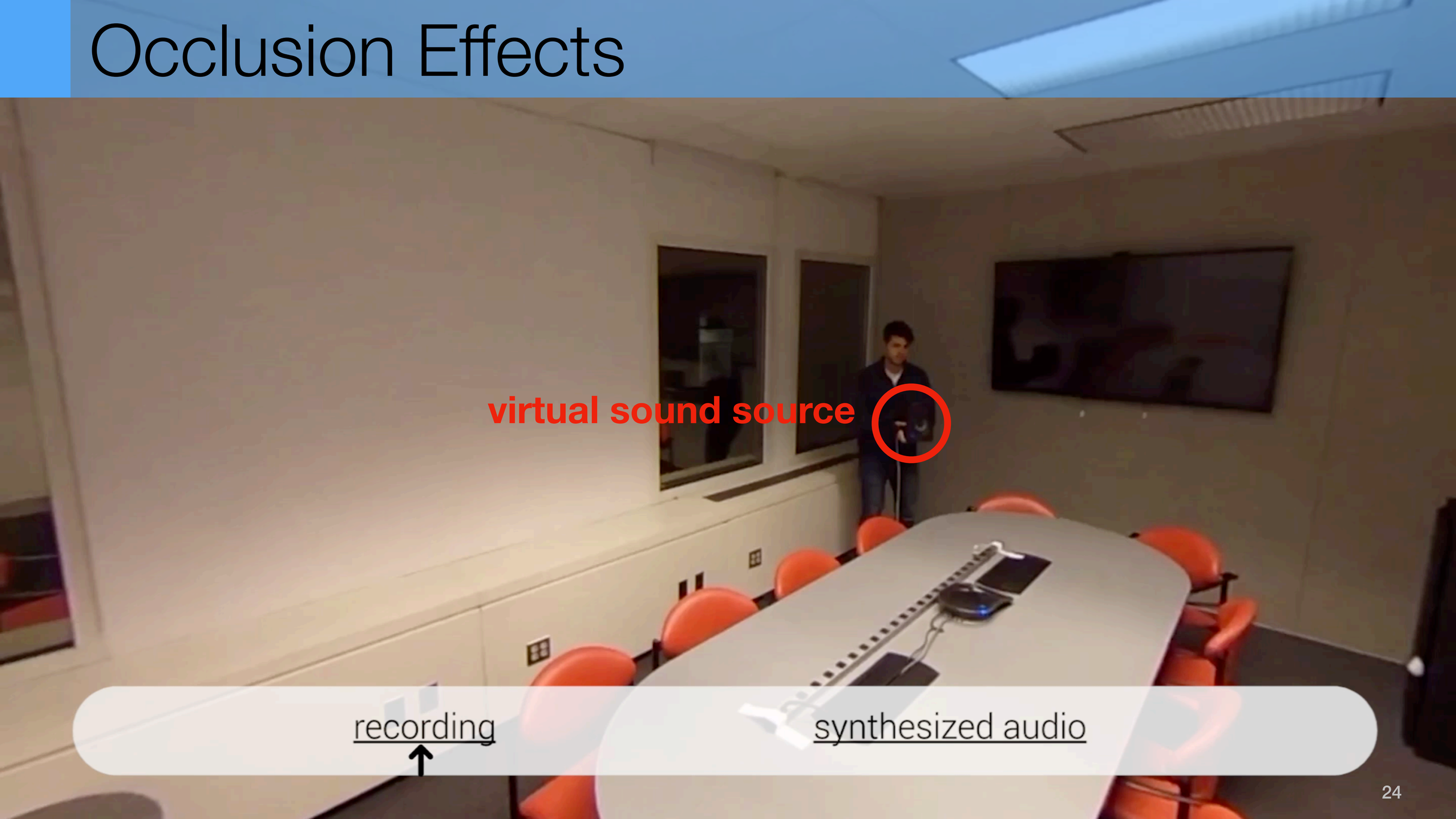
Noise Removal in Spatial Audio



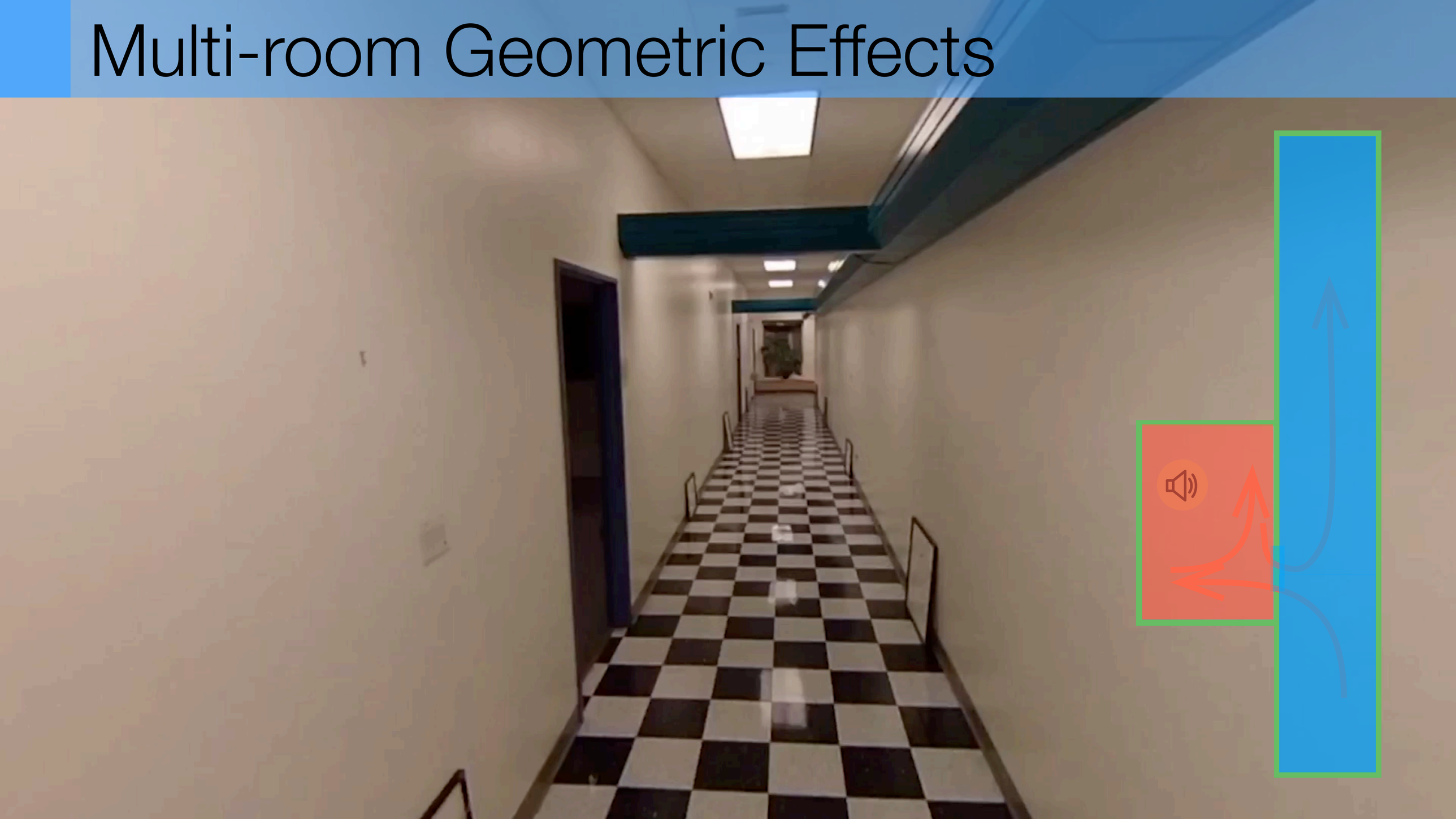
contaminated recording

clean synthesized audio

Occlusion Effects



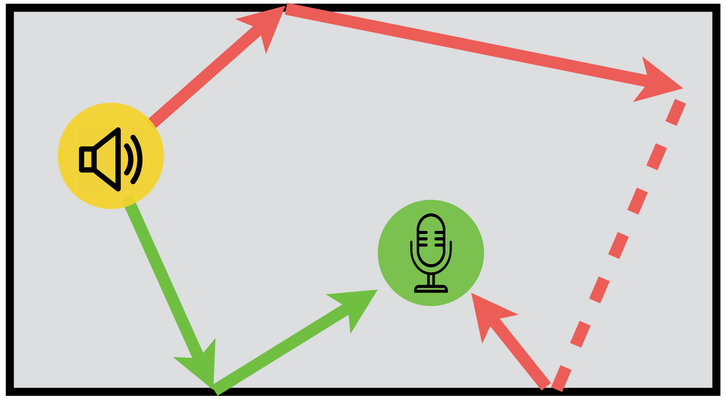
Multi-room Geometric Effects



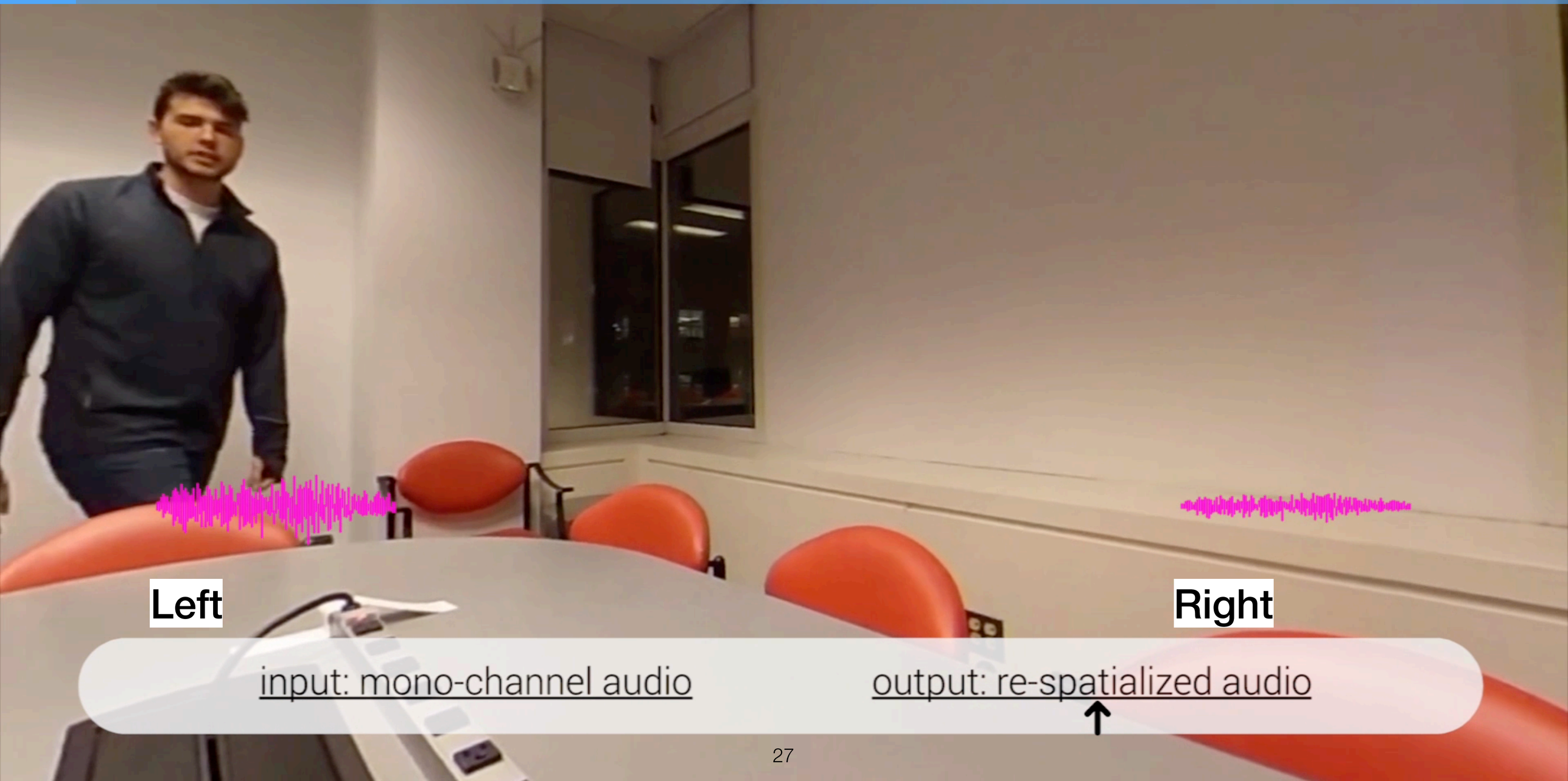
Re-spatialization of Mono-channel Audios

2D-3D conversion in movies

Re-spatialization in 360° videos



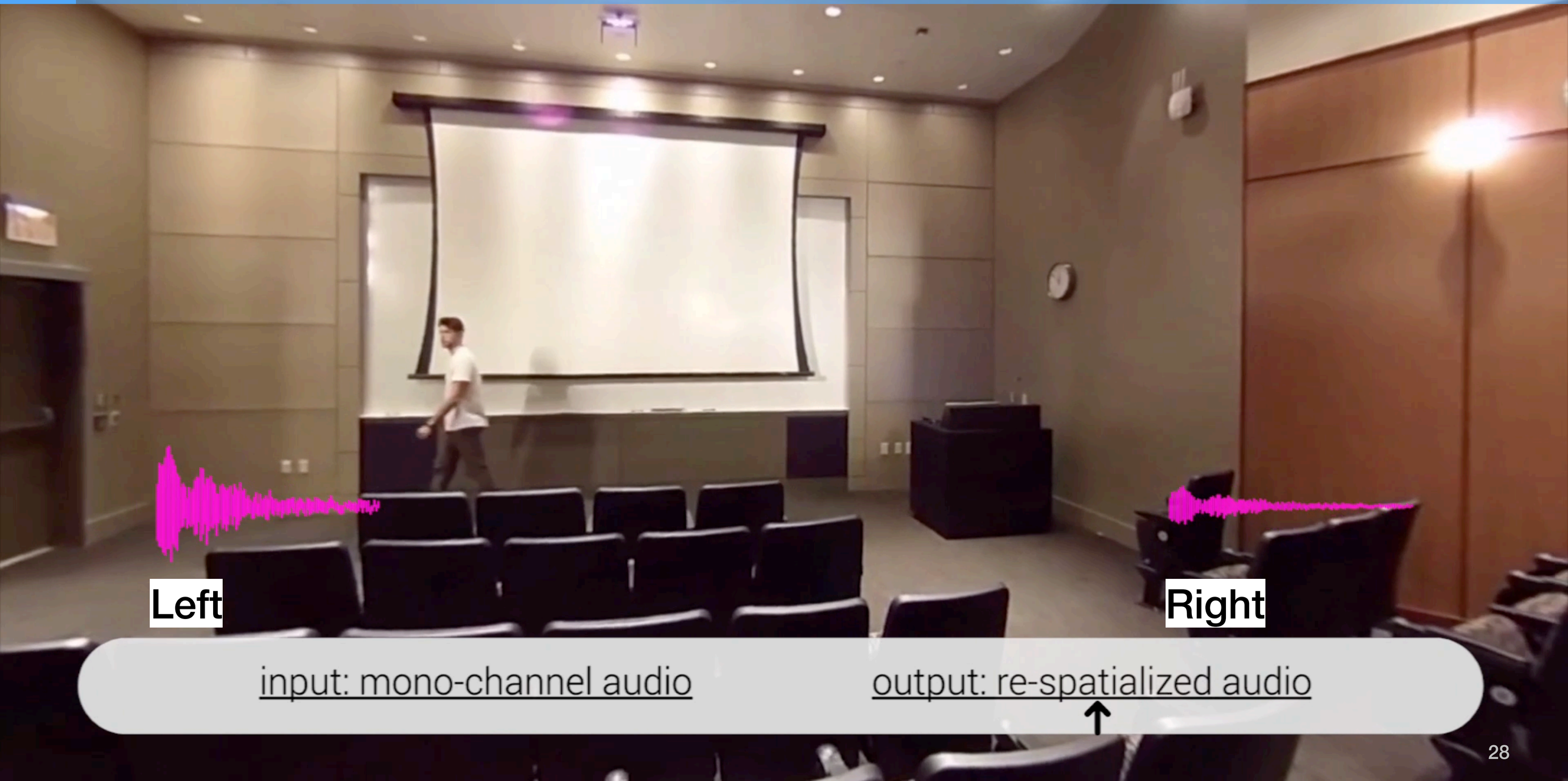
Re-spatialization



input: mono-channel audio

output: re-spatialized audio

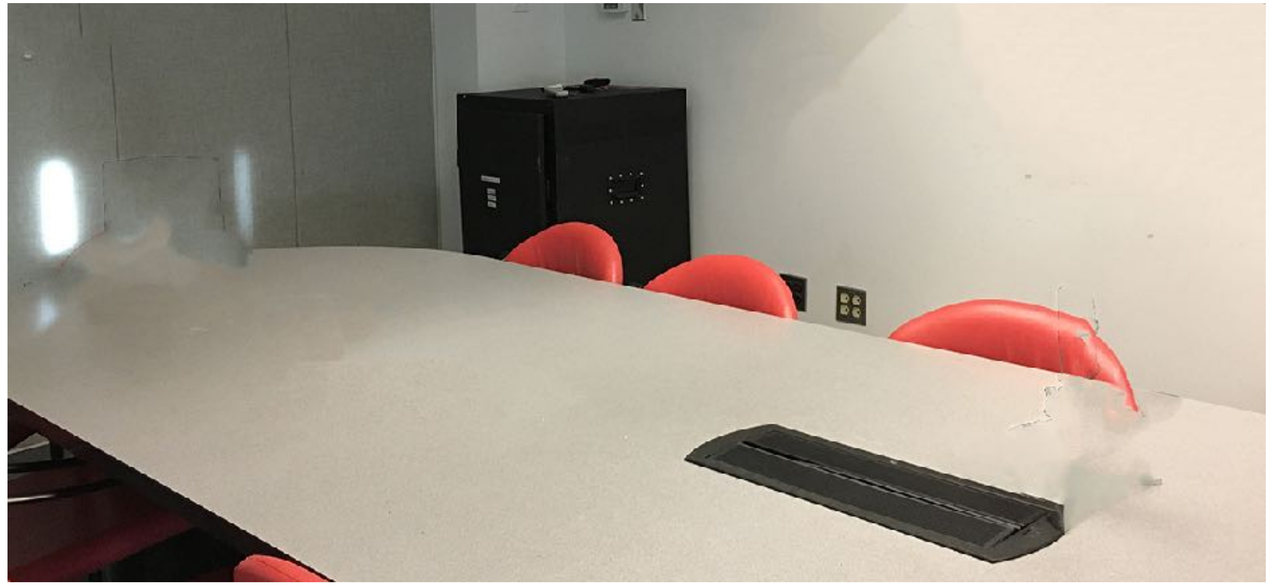
Re-spatialization



Conclusion

Scene-Aware Spatial Audio for 360° Videos

- matched the indoor scenes
- optimized simulation parameters
- efficient hybrid sound synthesis



Limitations and Future Work

- outdoor scenes
- distance to the listener/sources
- localized material estimation with multiple IRs
- passive audio feature retrieval

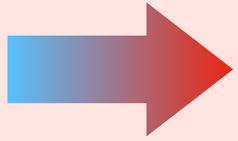
Acknowledgements

Chunxiao Cao, Zhili Chen, Carl Schissler, James Traer, Henrique Maia

sound tracing code **SfM code** **input audio data** **IR discussion** **proofread/voiceover**

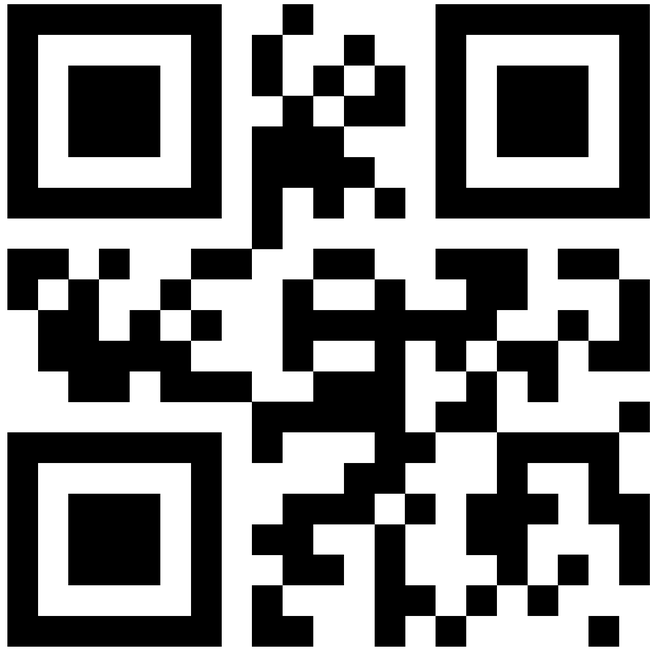
Adobe PhD Fellowship

Scene-Aware Audio for 360° Videos



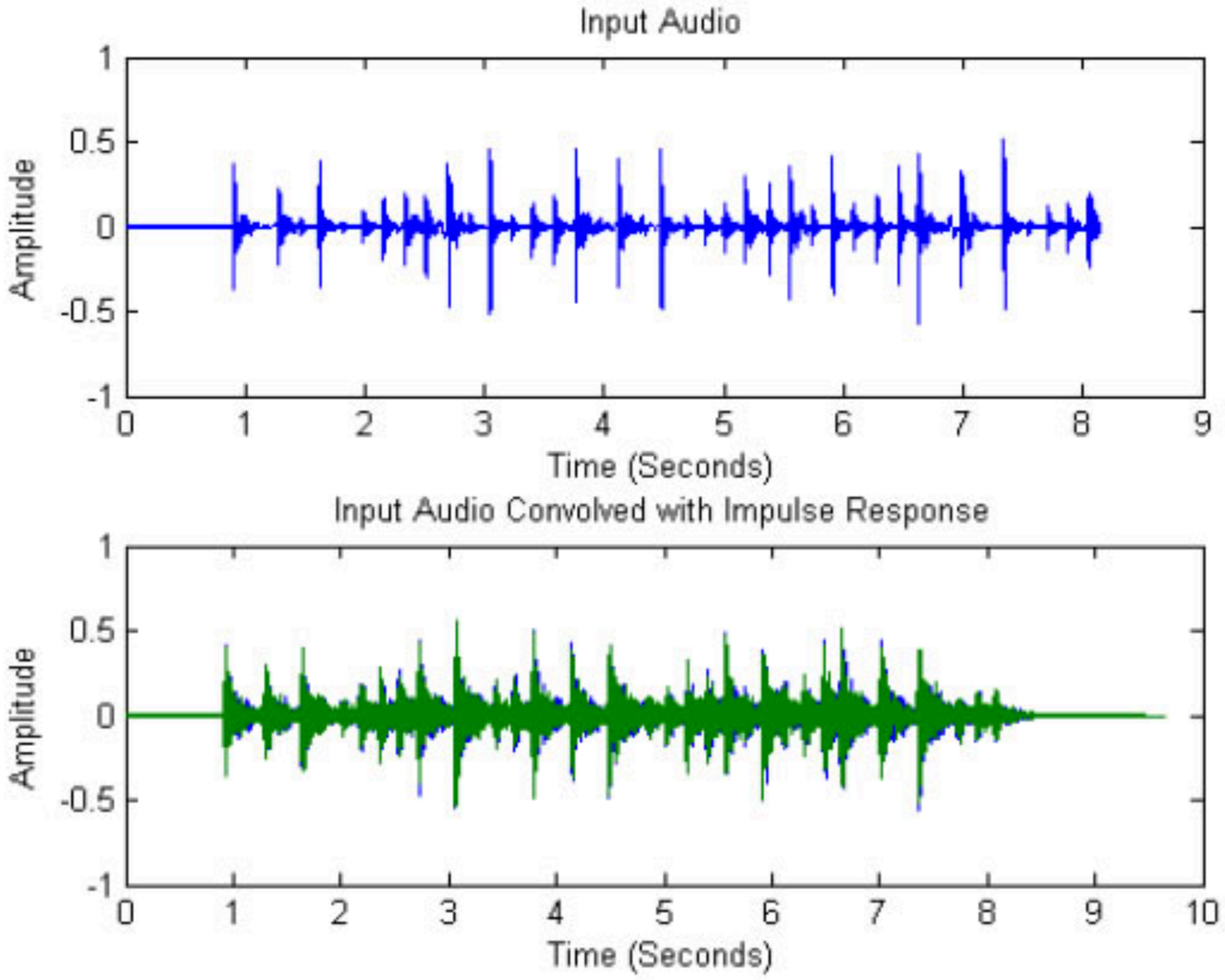
Dingzeyu Li
http://ding.fyi
Timothy R. Langlois
Changxi Zheng

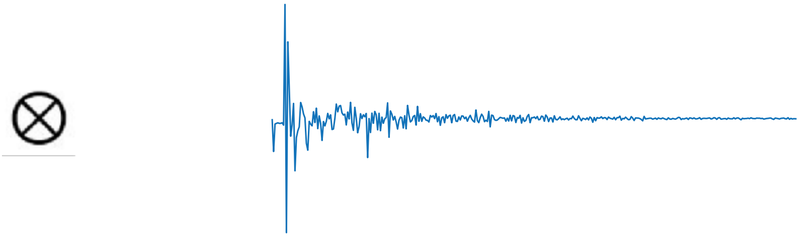
Thank you.



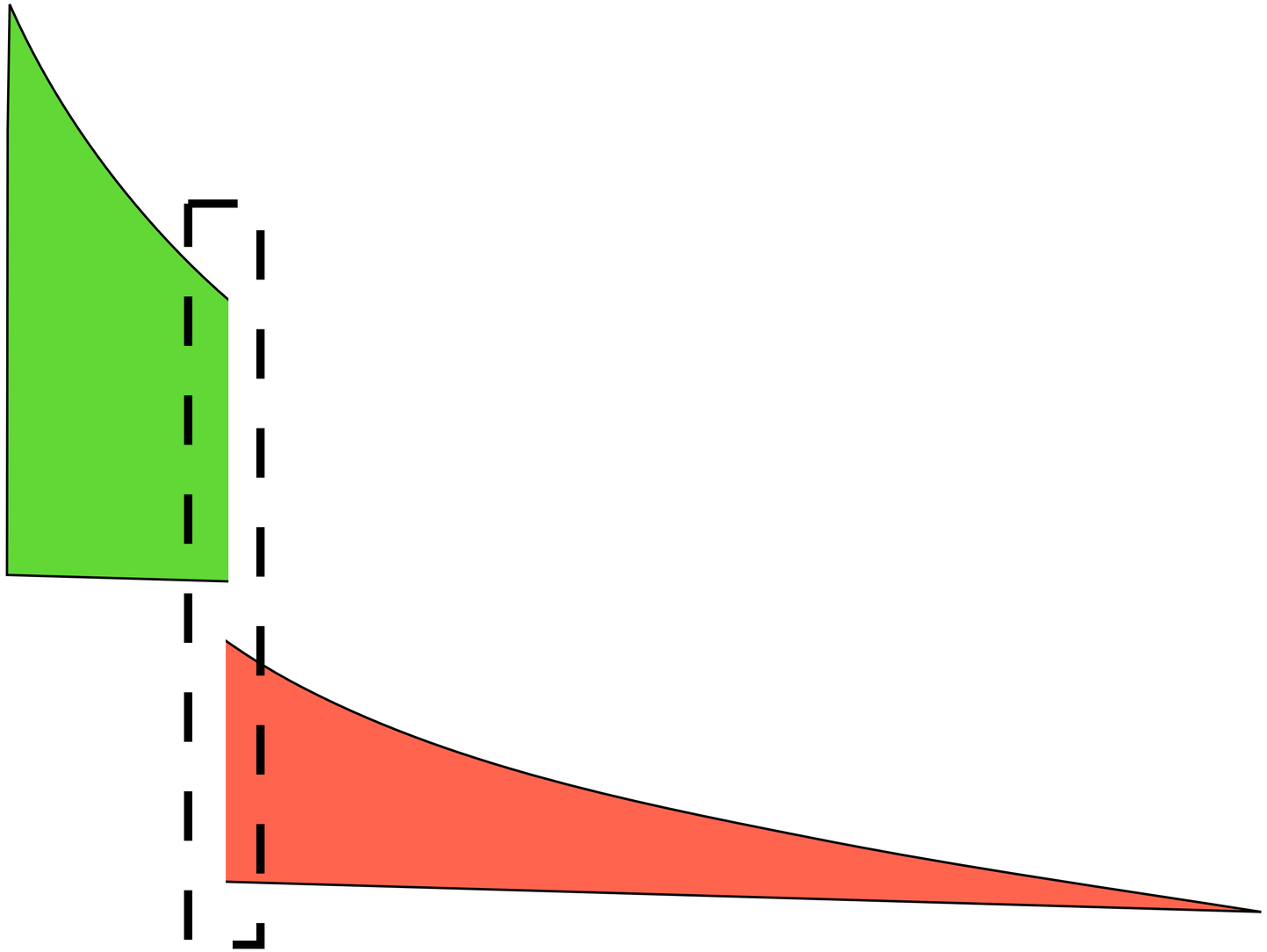
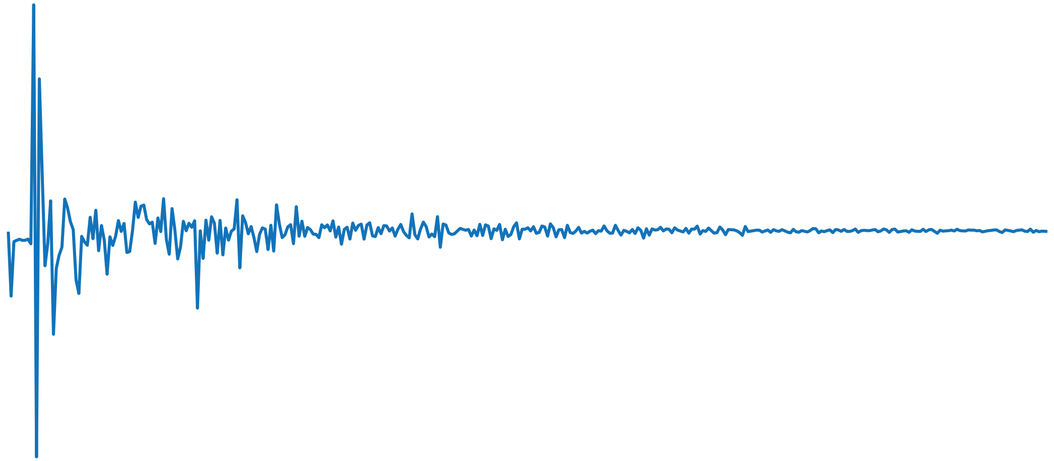
Additional Slides

Convolve with Impulse Response





Concatenation of Late Tail



Sample Impulse Response Clip