



# AV-GS: Learning material and geometry aware priors for Novel View Acoustic Synthesis Paper ID: 18643

**Swapnil Bhosale**  
University of Surrey, UK

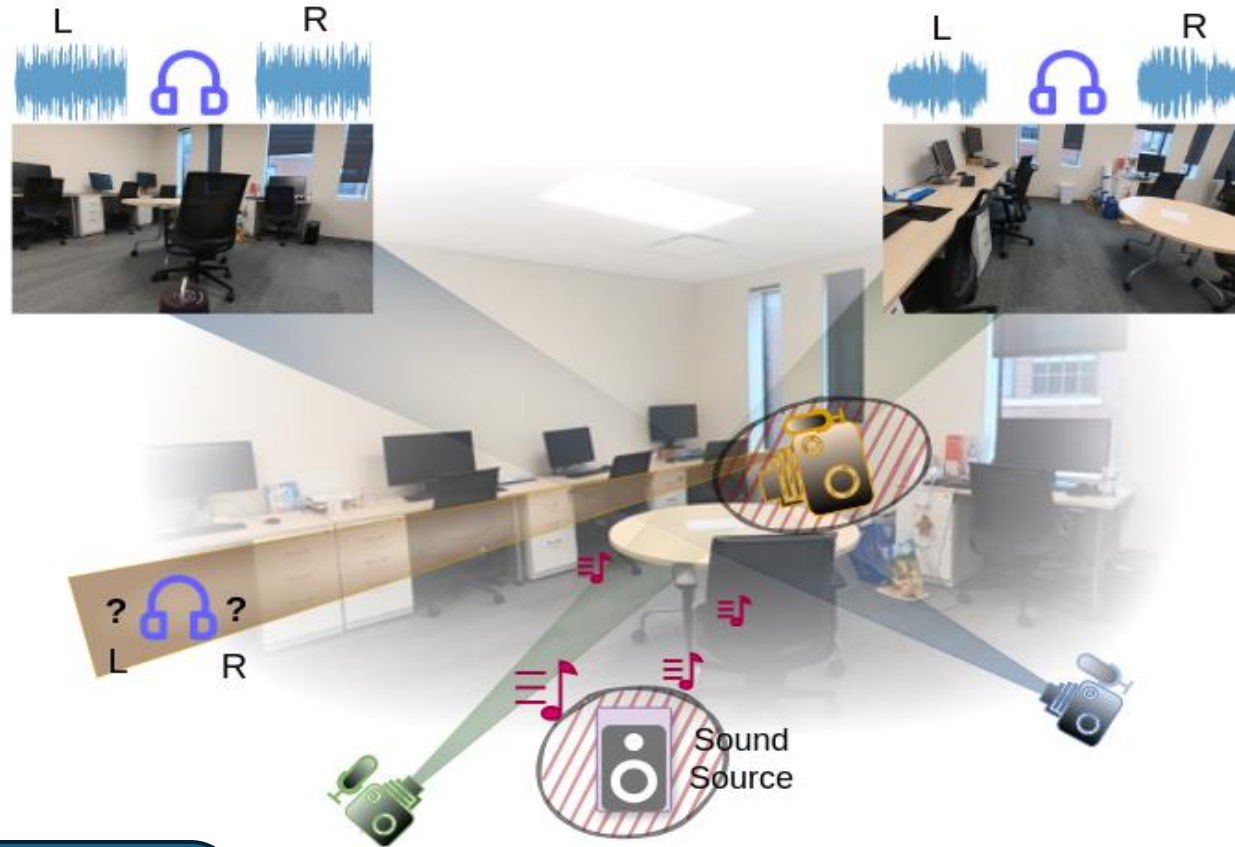
**Haosen Yang**  
University of Surrey, UK

**Diptesh Kanojia**  
University of Surrey, UK

**Jiankang Deng**  
Imperial College London, UK

**Xiatian Zhu**  
University of Surrey, UK

# Task - Novel view acoustic synthesis (NVAS)



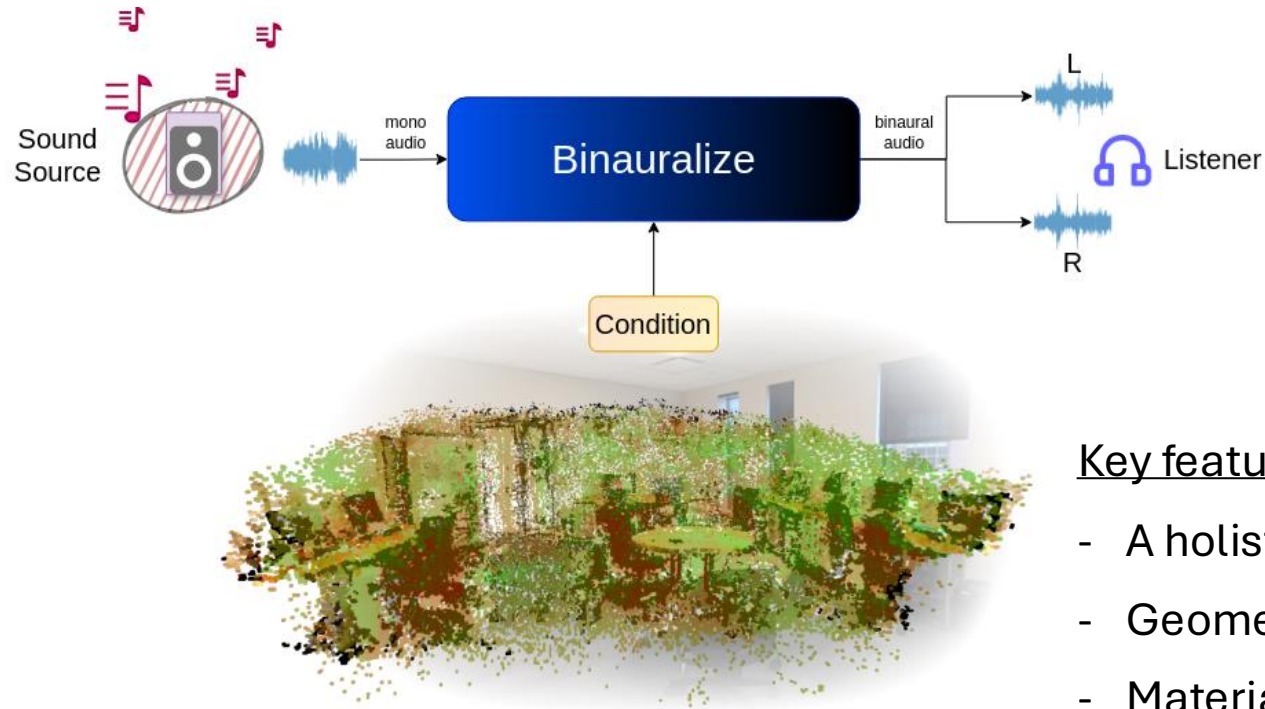
## *Given*

- (1) A sound source emitting mono audio within the 3D scene.
- (2) Binaural audios recorded from viewpoints distributed in the 3D scene.

## *Task*

Synthesize the binaural audio at an unseen viewpoint.

# Proposal – Audio-visual Gaussian Splatting (AV-GS)



## Key features:

- A holistic scene representation
- Geometry-awareness
- Material-awareness

# Proposal: AV-GS

## Stage-1

Learning locally initialized 3D Gaussians (G) using 3D Gaussian Splatting<sup>[1]</sup>.

## Decoupling physical geometry

Initializing audio-guidance parameters to obtain a *holistic geometry-aware material-aware* scene representation  $G_a$

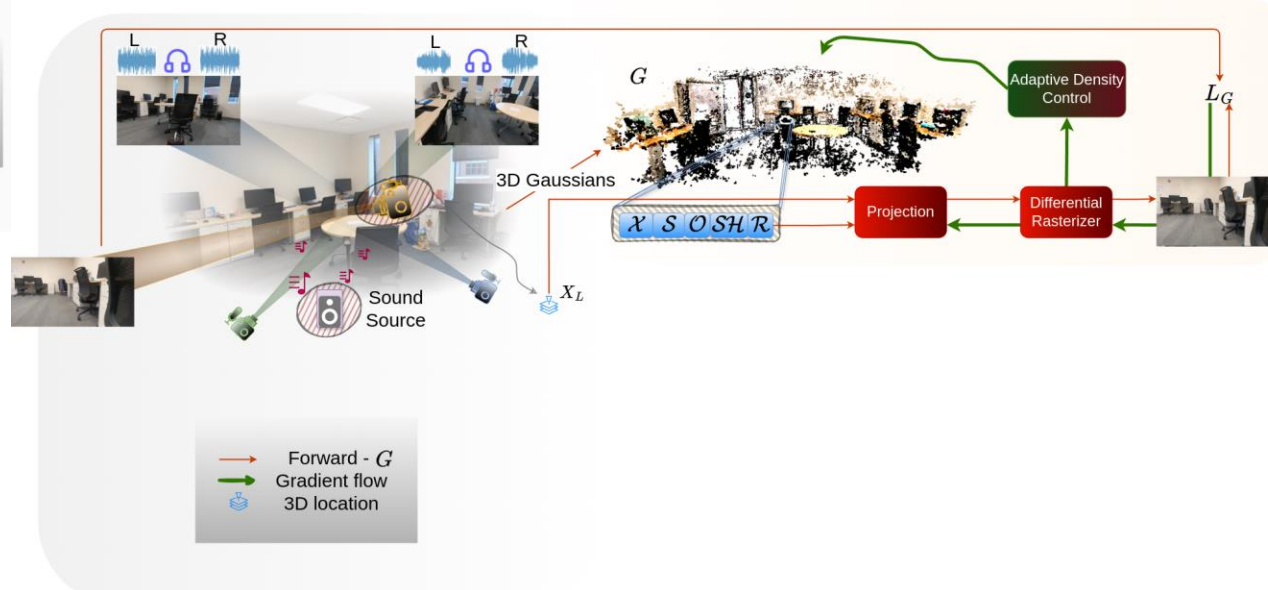
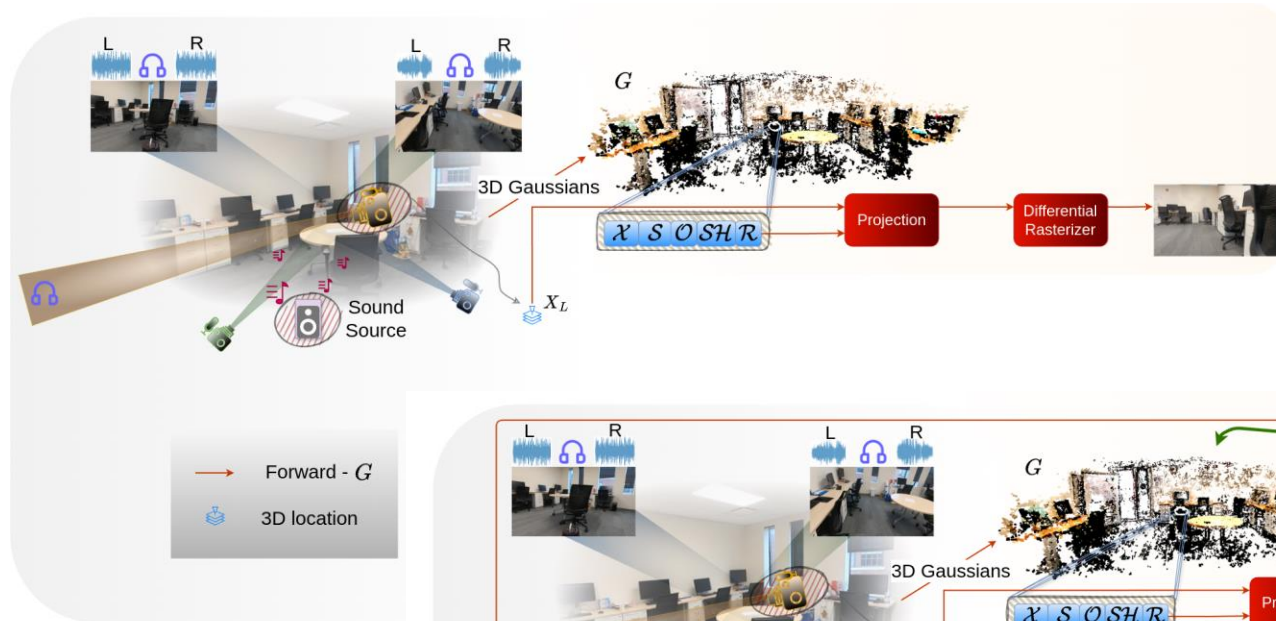
## Stage-2

Fusing audio-guidance parameters for points in the vicinity of the listener and sound source.

# Proposal: AV-GS

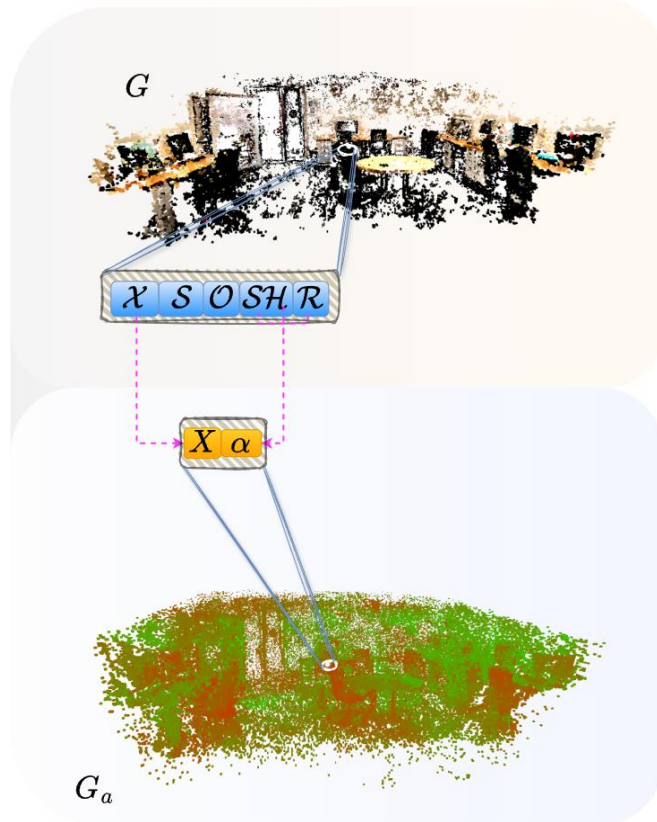
## Stage-1

Learning locally initialized 3D Gaussians ( $G$ ) using 3D Gaussian Splatting<sup>[1]</sup>.



[1] Kerbl, Bernhard, et al. "3d gaussian splatting for real-time radiance field rendering." ACM Transactions on Graphics (2023)

# Proposal: AV-GS



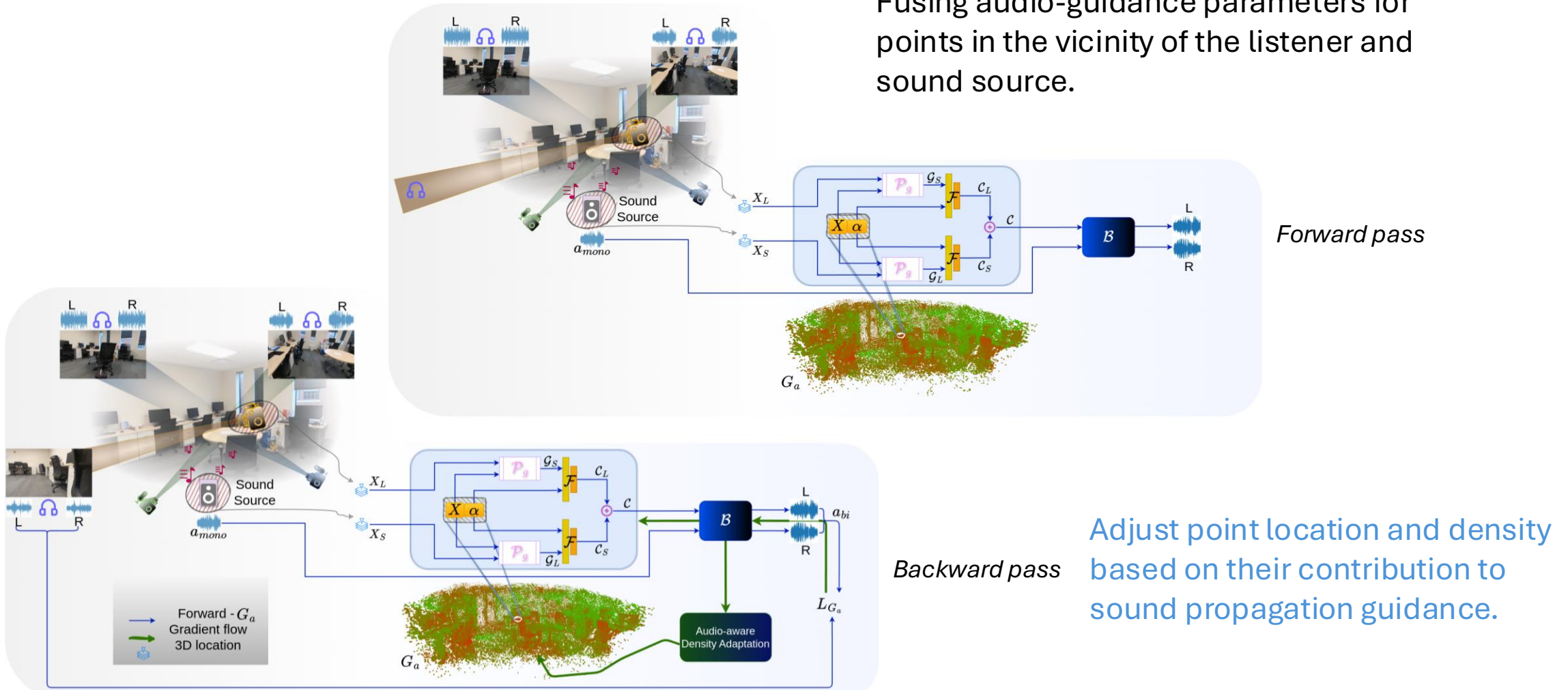
## Decoupling physical geometry

Initializing audio-guidance parameters to obtain a *holistic geometry-aware material-aware* scene representation  $G_a$

# Proposal: AV-GS

## Stage-2

Fusing audio-guidance parameters for points in the vicinity of the listener and sound source.



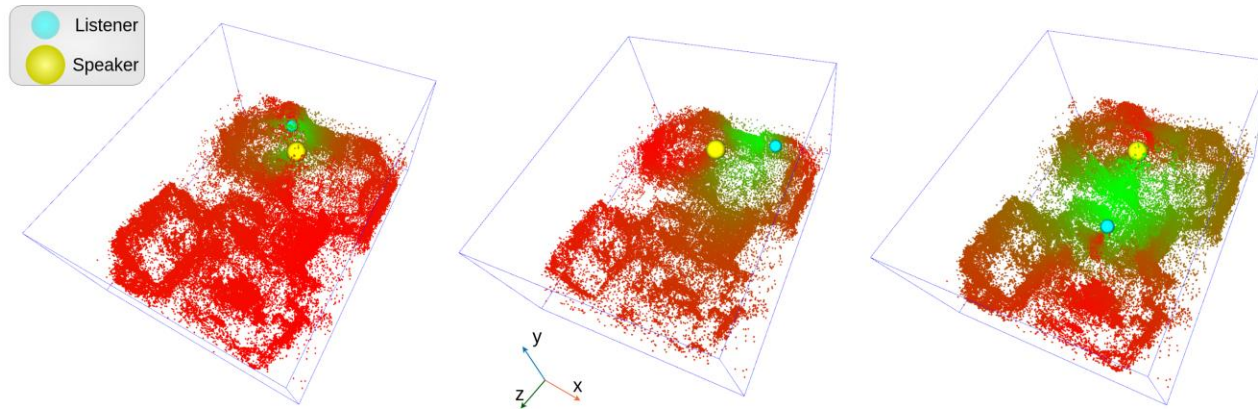
Forward pass

Backward pass

Adjust point location and density based on their contribution to sound propagation guidance.

# AV-GS: Highlights

## Explicit (+intuitive) modeling of Sound propagation



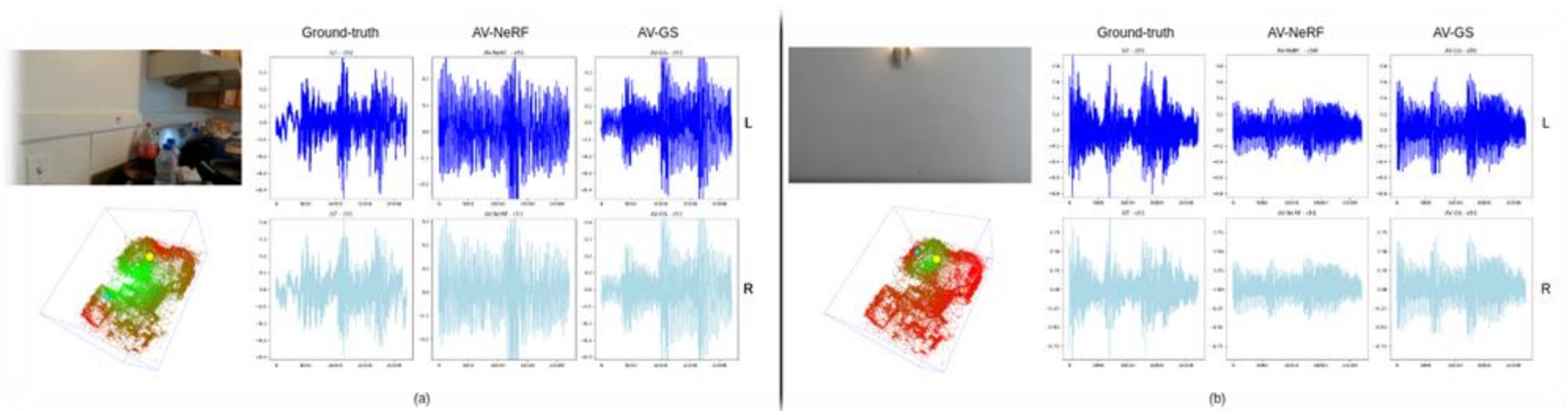
(We slice the scene into half along the  $y$ -axis, omitting the points from the ceiling, to facilitate better visibility.)

AV-GS learns an explicit point-based representation, with an audio-guidance feature w.r.t each point, that decides the contribution of the point in providing audio guidance for a given listener and sound source pair. Notice how the points behind the speaker or points behind rigid walls do not contribute in guidance to sound propagation, and hence have lower values (red).



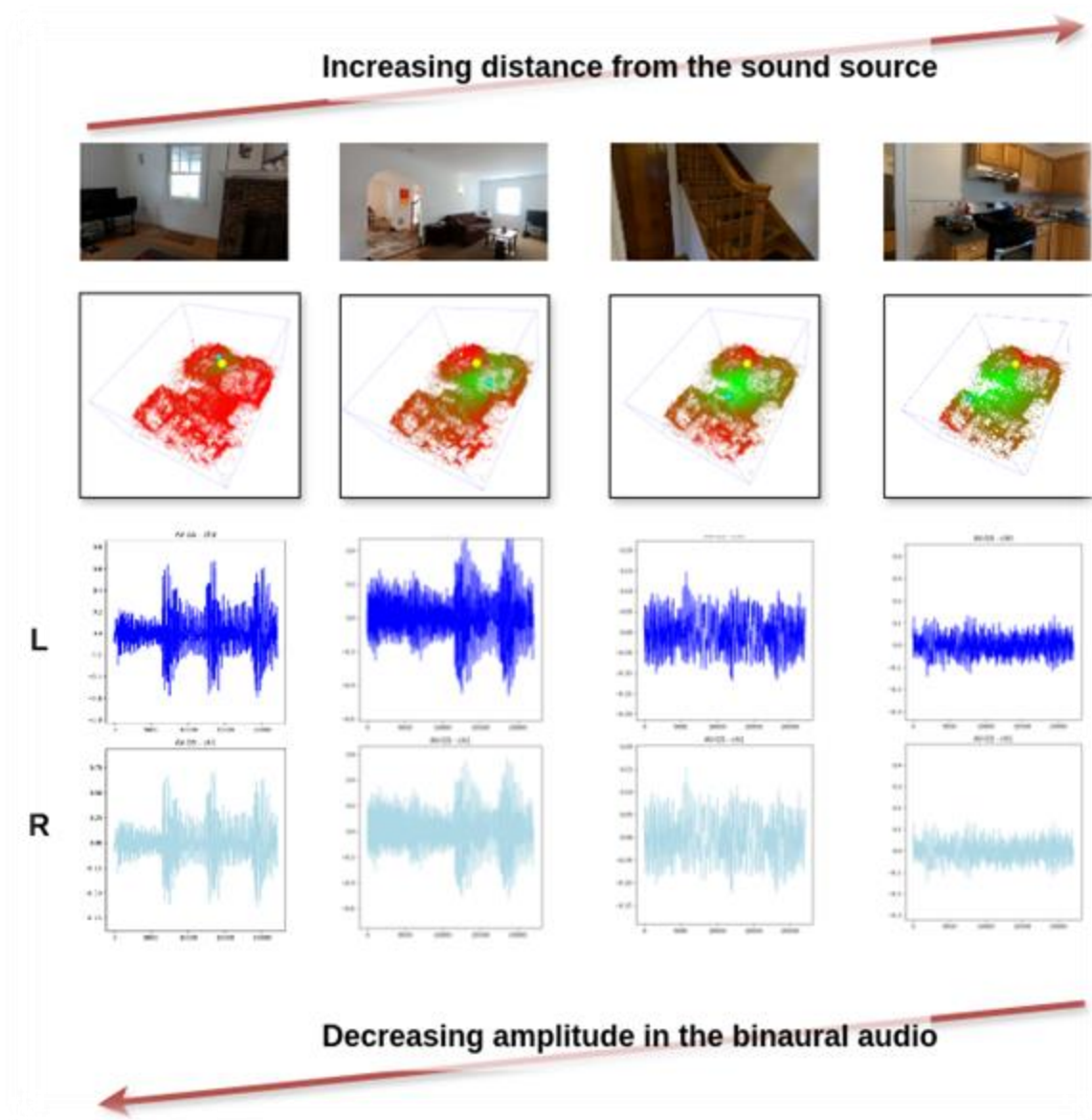
# AV-GS: Highlights

## Informative conditioning



In the presence of (a) complex geometry (left), and (b) meaningless views (right), AV-NeRF makes errors in binaural synthesis. The learned holistic scene representation, used by AV-GS, is unaffected by both scenarios and generates informative conditions.

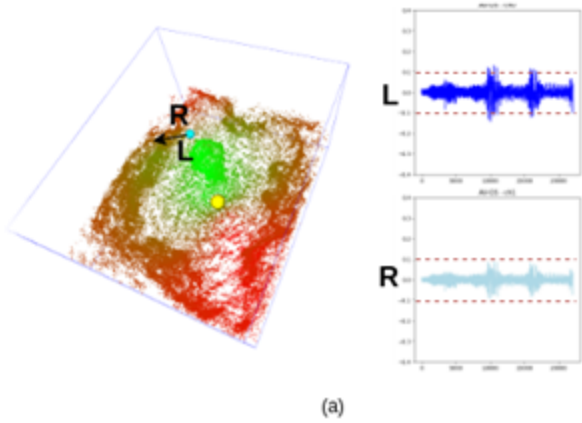
# AV-GS: Highlights



**Distance aware audio rendering**  
As the distance from the sound source increases the amplitude of the synthesized binaural audio decreases (exhibiting doppler's effect).

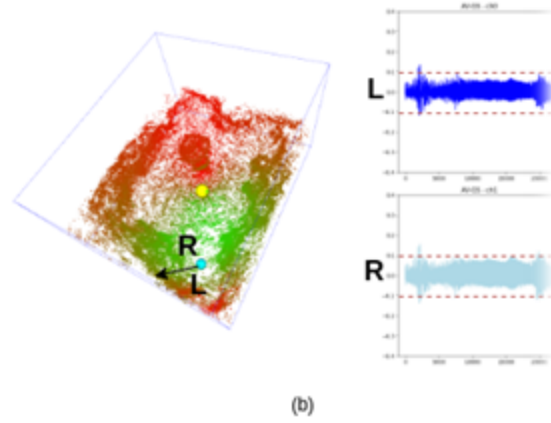
# AV-GS: Highlights

Sound source on the left side of listener



Higher amplitude of audio on the synthesized left channel.

Sound source on the right side of listener



Higher amplitude of audio on the synthesized right channel.

## **Direction-aware audio rendering**

Relative to the viewing direction of the listener w.r.t to the sound source, AV-GS synthesizes varying amplitude levels in the left and right audio channel.



Thank you!

AV-GS:  
Learning material and geometry aware priors for  
Novel View Acoustic Synthesis  
Paper ID: 18643

**Swapnil Bhosale**  
University of Surrey, UK

**Haosen Yang**  
University of Surrey, UK

**Diptesh Kanojia**  
University of Surrey, UK

**Jiankang Deng**  
Imperial College London, UK

**Xiatian Zhu**  
University of Surrey, UK