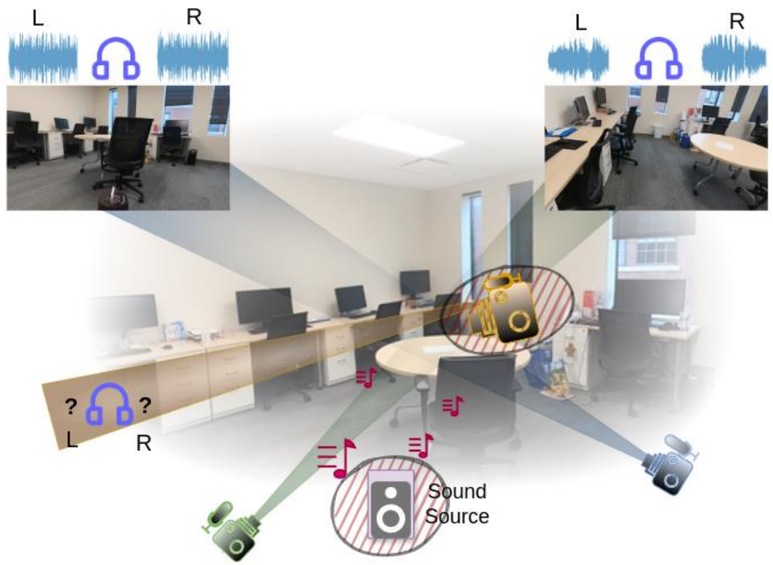
AV-GS: Learning material and geometry aware priors for Novel View Acoustic Synthesis

Swapnil Bhosale[1], Haosen Yang[1], Diptesh Kanojia[1], Jiankang Deng[2], Xiatian Zhu[1]

[1] University of Surrey, UK
[2] Imperial College London, UK

Task: Novel View Acoustic Synthesis



Given

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

Task

Synthesize the binaural audio at an unseen viewpoint.

Motivation: Binauralization with holistic conditioning

Incorporate comprehensive contextual knowledge (a holistic scene representation) beyond the listener's field of view for an enhanced conditioning.

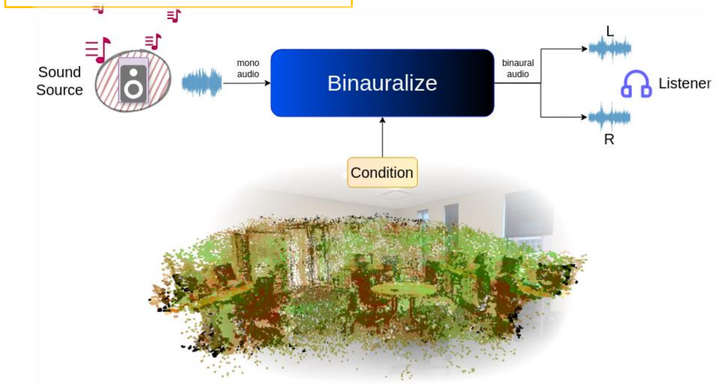
3D-GS already learns comprehensive scene representations, adopt directly for binauralization?

3D-GS point optimization driven by view-reconstruction loss (purely image domain):
- **Over-populate along object edges**
- **Under-populate along texture-less regions (walls, doors etc.)**

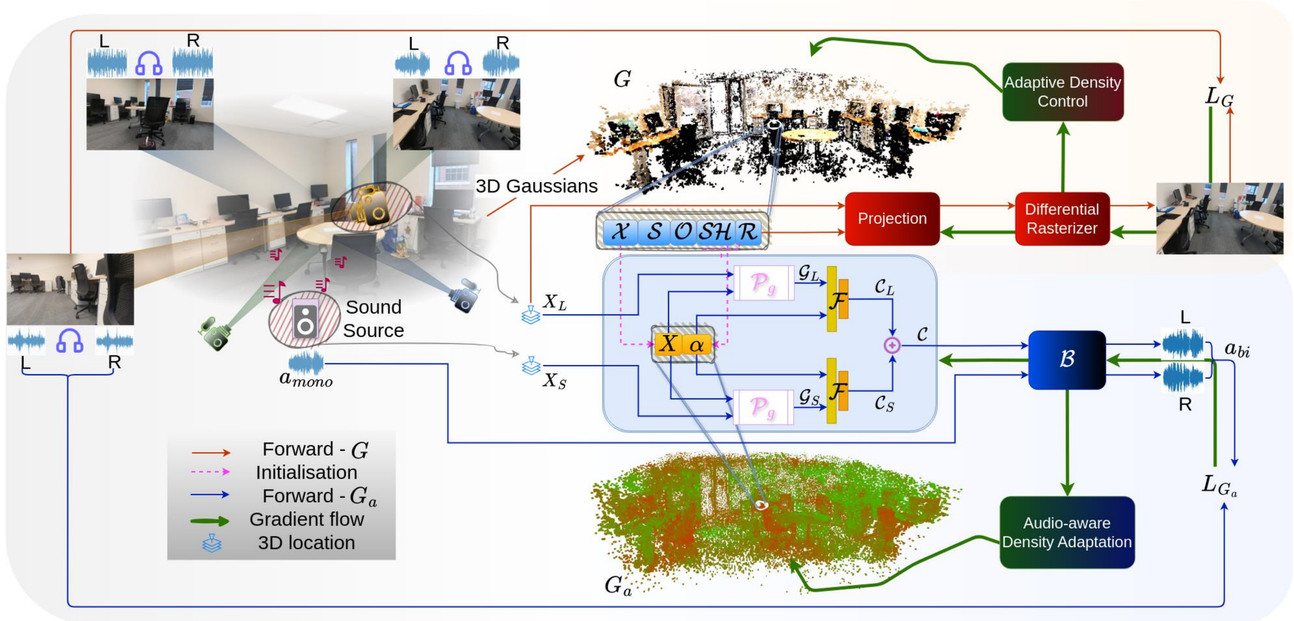
Rhetoric for learning sound propagation :(
major changes in sound paths (absorption and diversion) primarily happen around those texture-less regions.

Realistic binaural audios need,
- **Direction-awareness**
- **Distance-awareness**

Jointly model 3D geometry and material characteristics of the visual scene objects.



PROPOSAL



Stage-1

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

Decoupling physical geometry

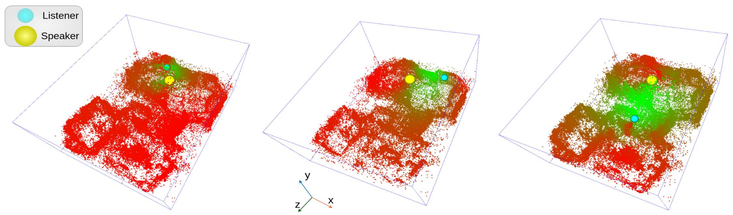
Initializing audio-guidance parameters to obtain a *holistic geometry-aware material-aware* scene representation G_a

Stage-2

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

Highlights:

(i) Intuitive modeling of sound propagation

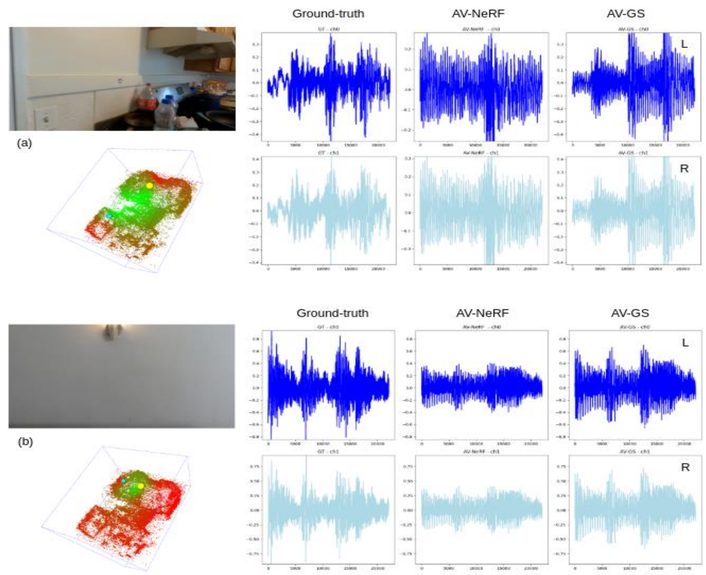


(We slice the scene into half along the y-axis, omitting the points from the ceiling, to facilitate better visibility.)

Points learn audio-guidance features for each listener-source pair. Points behind the speaker or rigid walls show lower values (red), indicating minimal contribution to sound propagation guidance.

(ii) Informative conditioning

In complex geometry (top) and meaningless views (bottom), errors occur in binaural synthesis in prior art[2].
However, the holistic scene representation remains unaffected and provides informative conditions.



References

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

[2] Liang, Susan, et al. "Av-nerf: Learning neural fields for real-world audio-visual scene synthesis." NeurIPS 2023.