

AV-Cloud: Spatial Audio Rendering Through Audio-Visual Cloud Splatting

Mingfei Chen, Eli Shlizerman
University of Washington



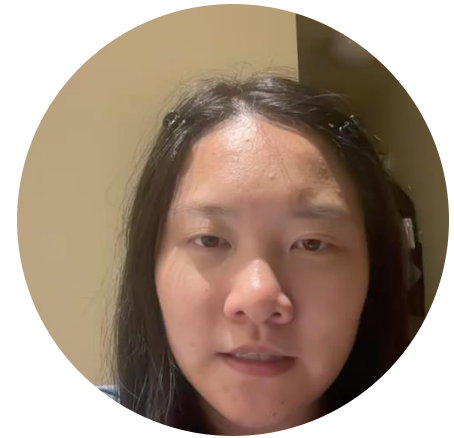
Please turn the Audio ON



NeuroAI Lab:

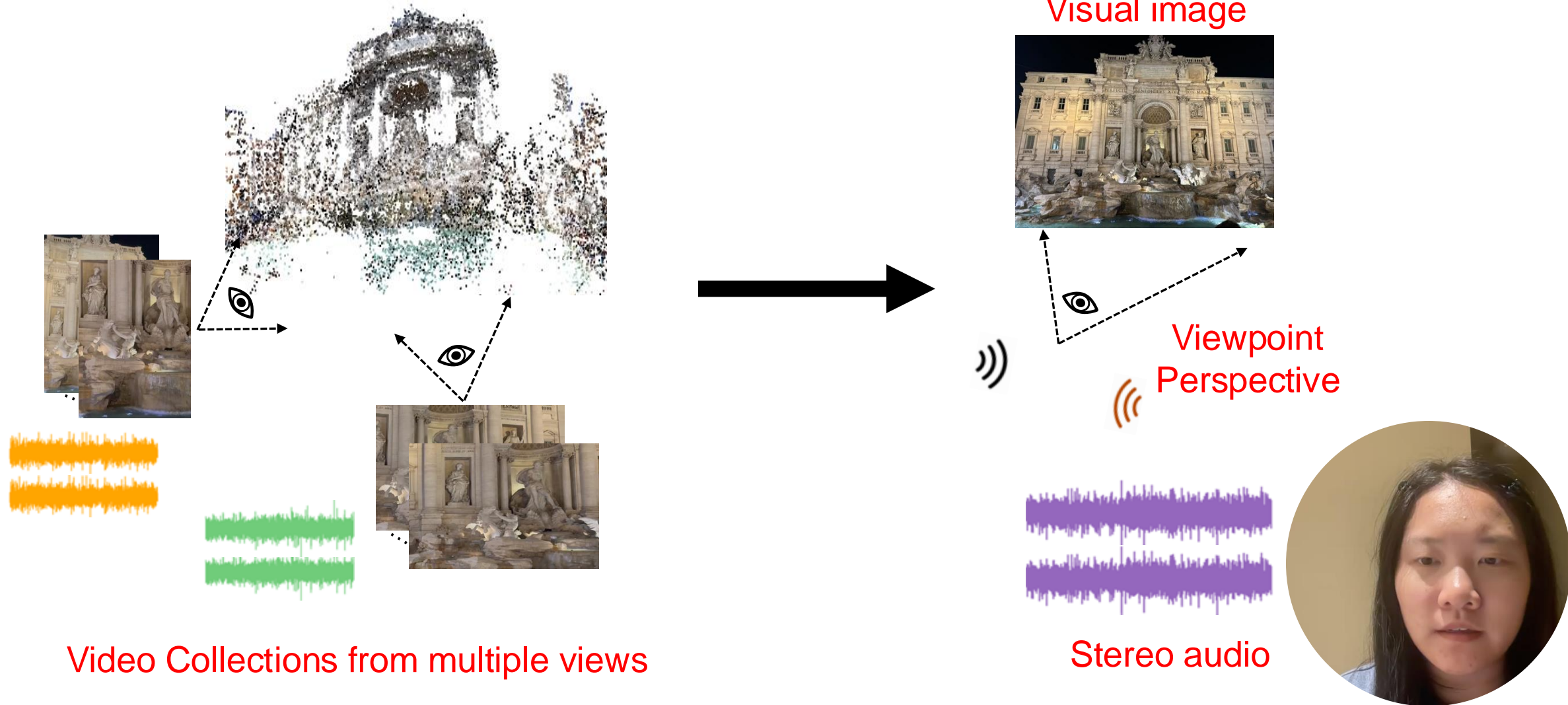


Spatial Audio Reconstruction of Audio-Visual Scenes with **In-the-Wild Videos**

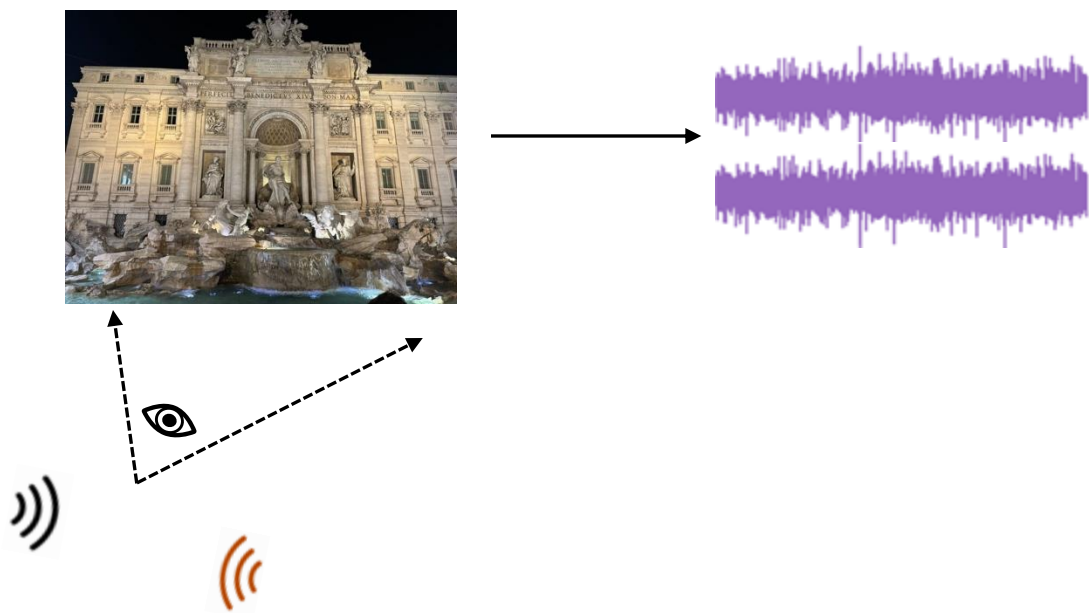


Drumheller Fountain @ University of Washington, Seattle

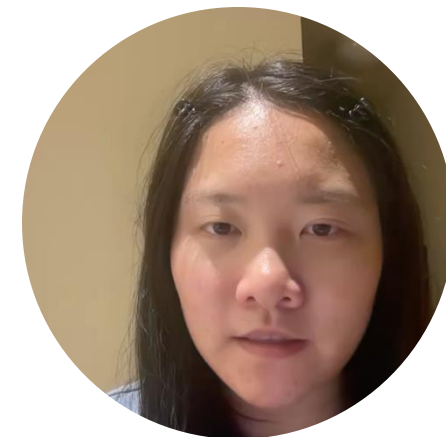
3D Audio-Visual Rendering



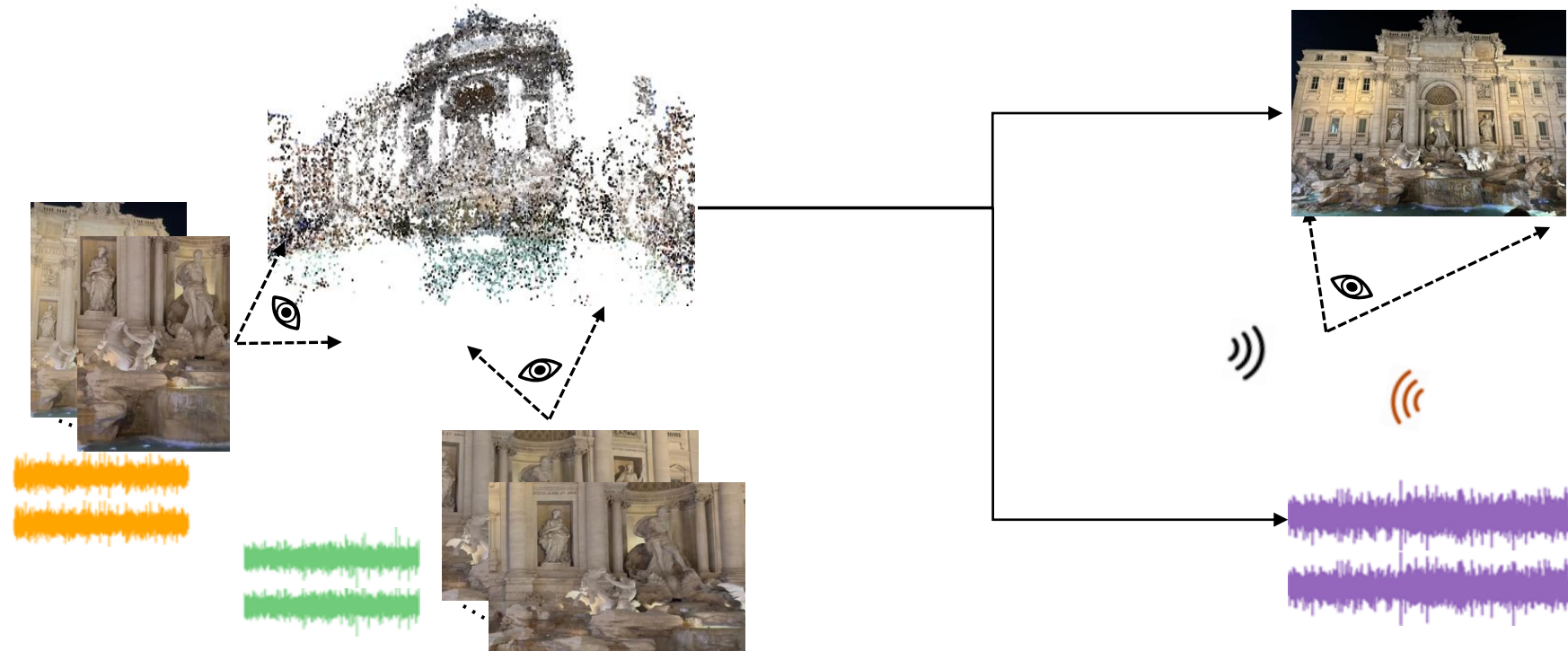
Existed Works: Visually-Guided Spatial Audio Render



- Depends on visual render quality
- Time delay for audio render



AV-Cloud can synchronize with visual rendering



Without pre-rendered images or explicit emitter locations, AV-Cloud:

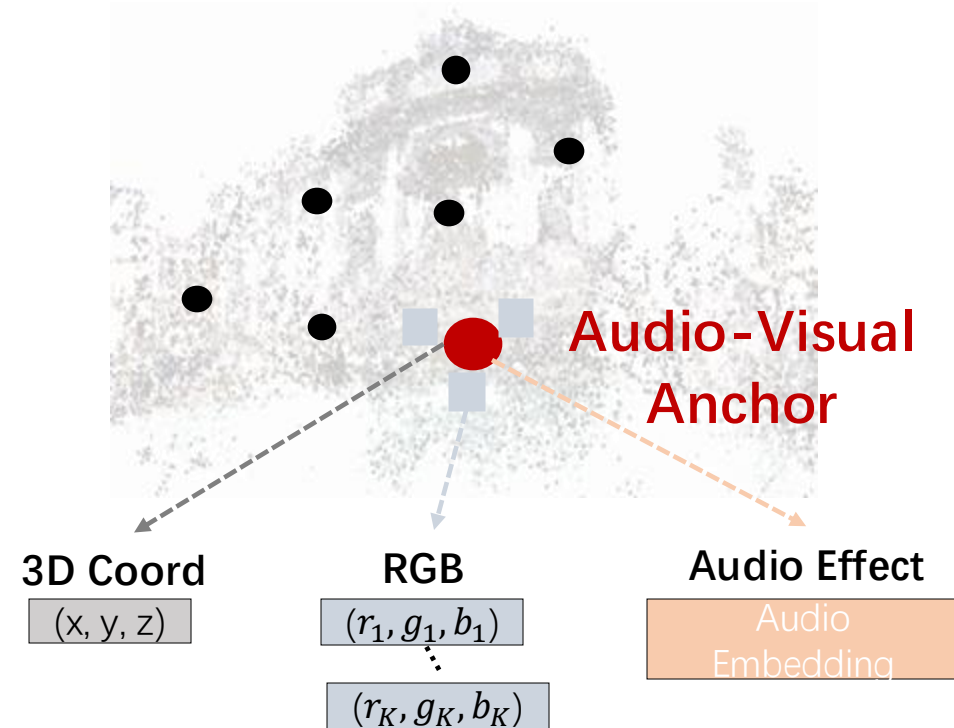
- 1) Audio rendering directly from scene representations
- 2) Outperforms previous methods in accuracy
- 3) Real-time rendering speed



How does *AV-Cloud* achieve that?

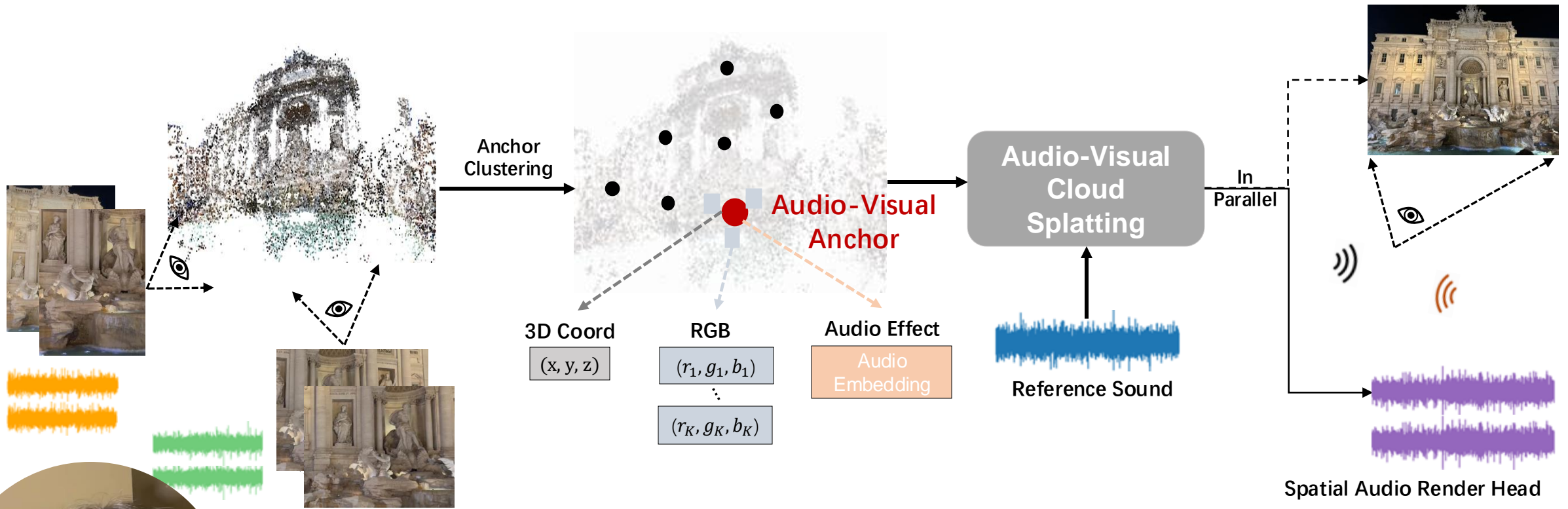


AV-Cloud: Effective Audio-Visual Representation for 3D Scenes



AV-Cloud constructs **Audio-Visual Anchors** for scene representation and transforms **monaural reference** sound into **spatial audio**.

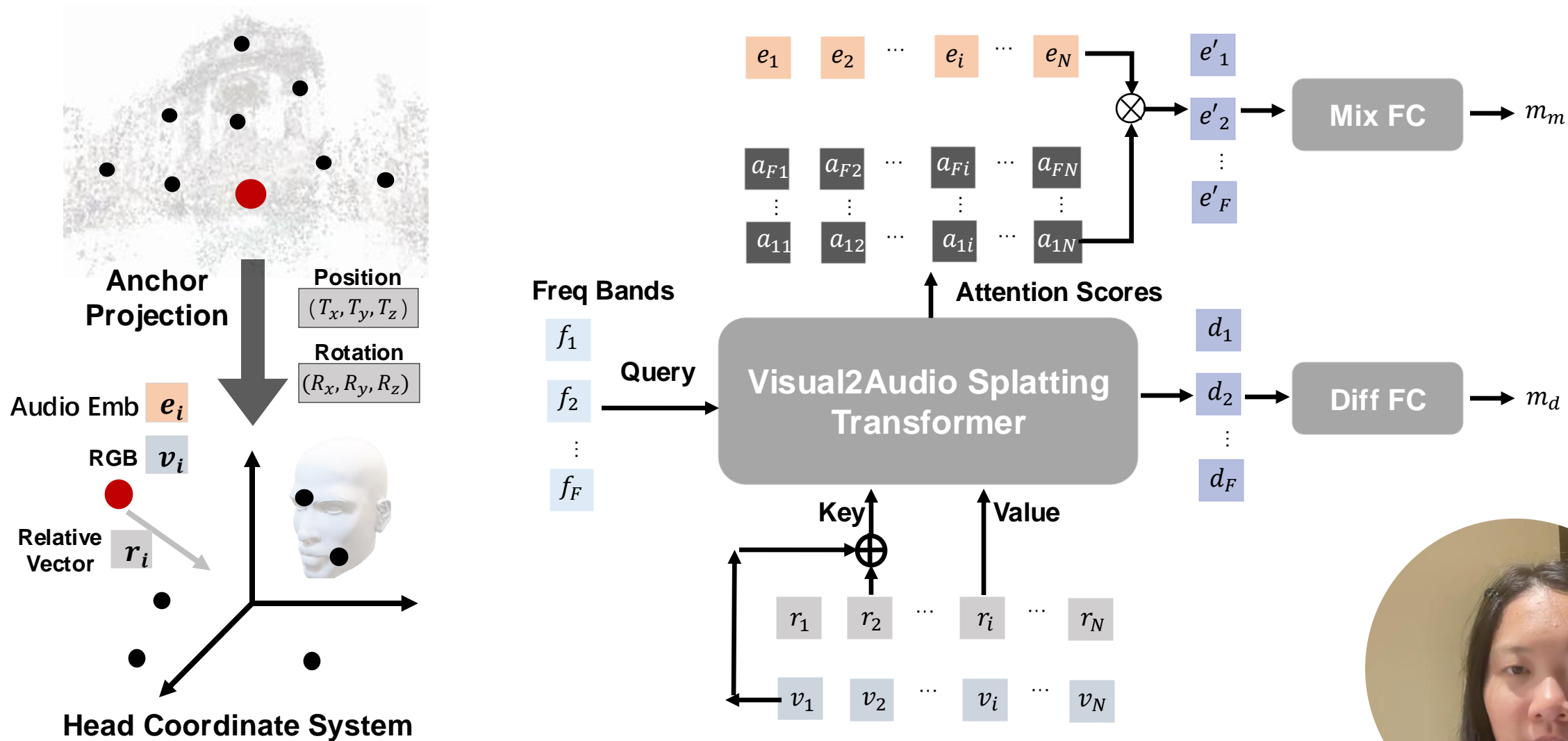
Overview of Audio-Visual Cloud Splatting (AVCS)



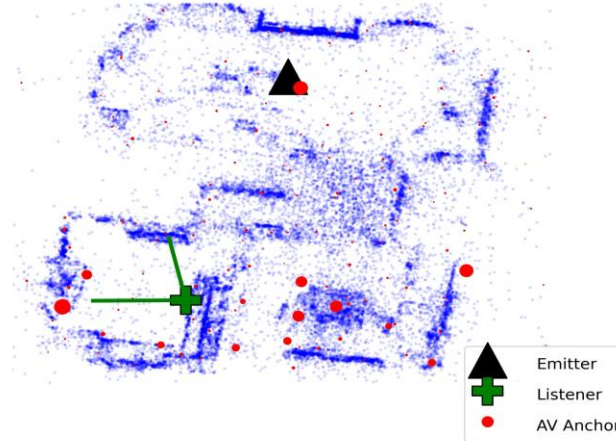
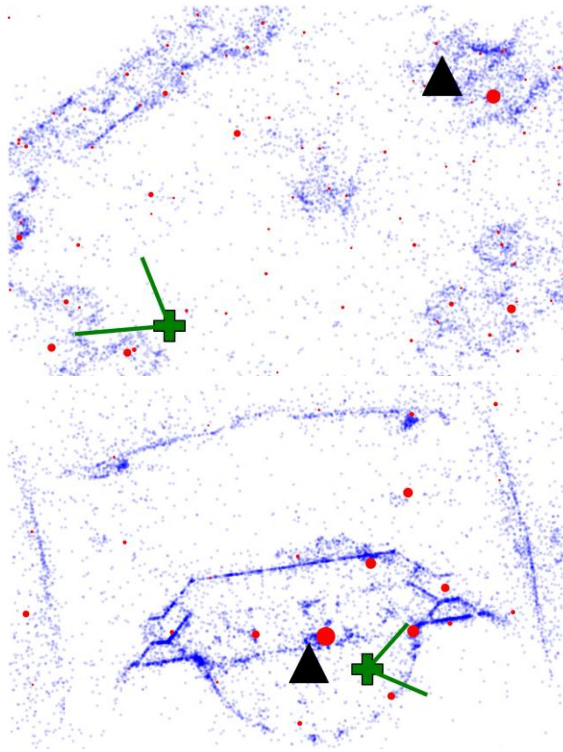
How does **AVCS** render spatial audio
from AV-Cloud?



Audio-Visual Cloud Splatting



AV-Cloud Attention Weights Visualization



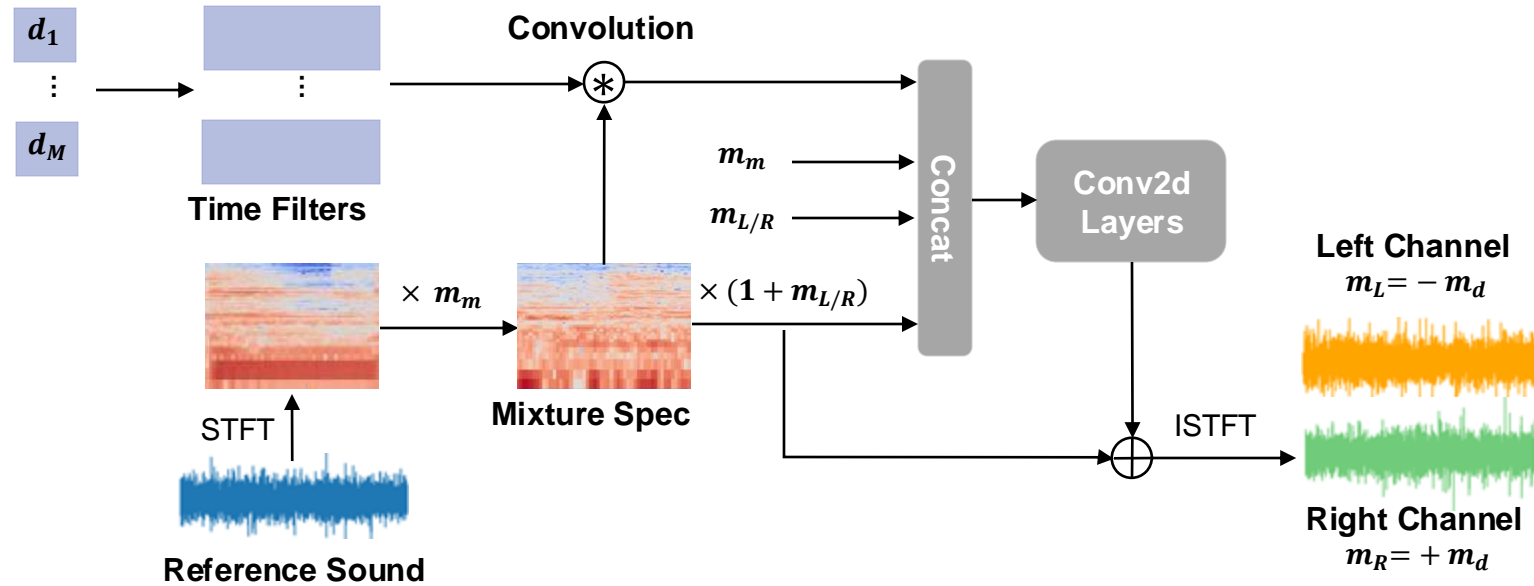
Higher attention values on anchors:

- 1) On key boundaries
- 2) Close to listener location
- 3) Close to the emitter location

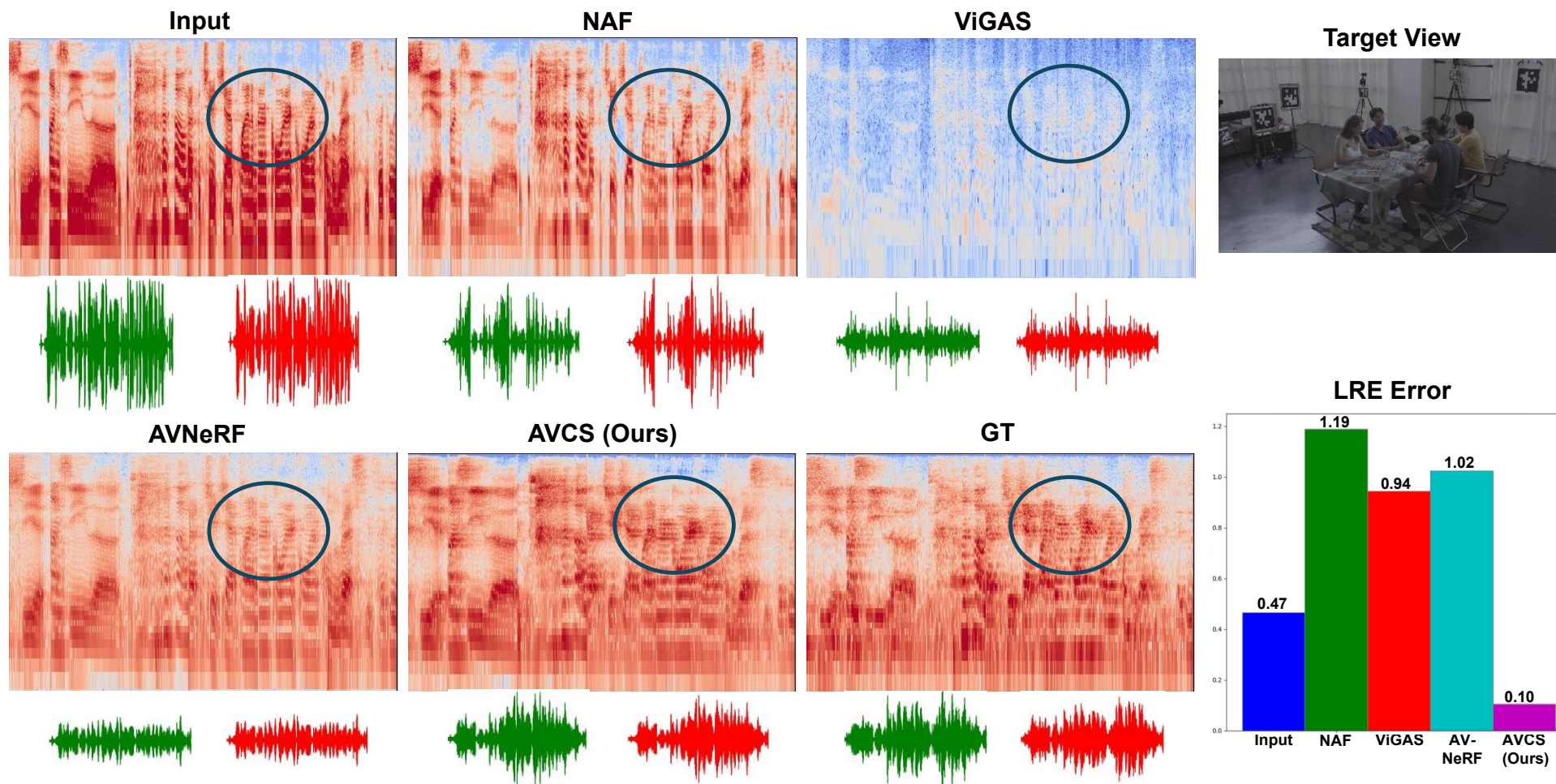


Spatial Audio Render Head

Adjust audio energy on
the time domain



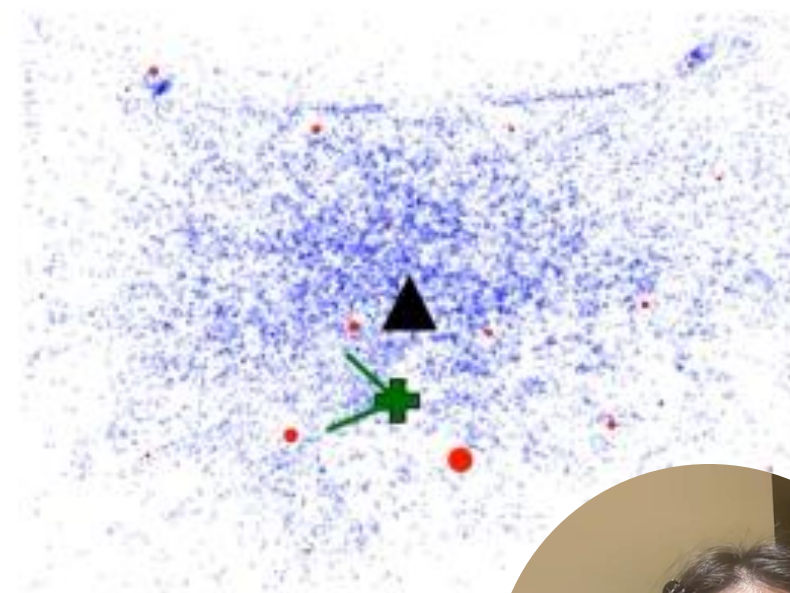
Qualitative Results Comparison



25 FPS Audio-Visual Rendering on Mac M2



Walk around the black speaker in the park



Thank you for watching

**AV-Cloud: Spatial Audio Rendering
Through Audio-Visual Cloud Splatting**

NeurIPS 2024

