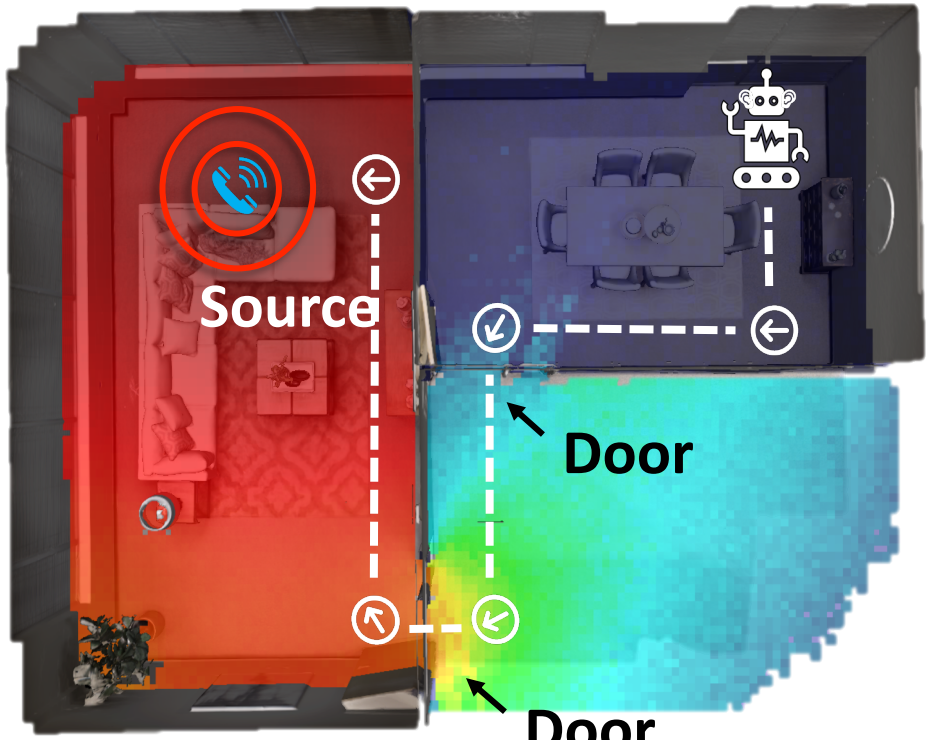
# SoundSpaces: Audio-Visual Navigation in 3D Environments

*Changan Chen*[*1,4], *Unnat Jain*[*2,4], *Carl Schissler*[3], *Sebastia V. Amengual Gari*[3], *Ziad Al-Halah*[1], *Vamsi K. Ithapu*[3], *Philip Robinson*[3], *Kristen Grauman*[1,4]

[1]UT Austin, [2]UIUC, [3]Facebook Reality Labs, [4]Facebook AI Research




ECCV'20
SEC, Glasgow
23-28 AUGUST 2020

# Navigation Is a Multisensory Experience

We often use *vision*, *audio*, *touch*, *smell* to move around in the environment

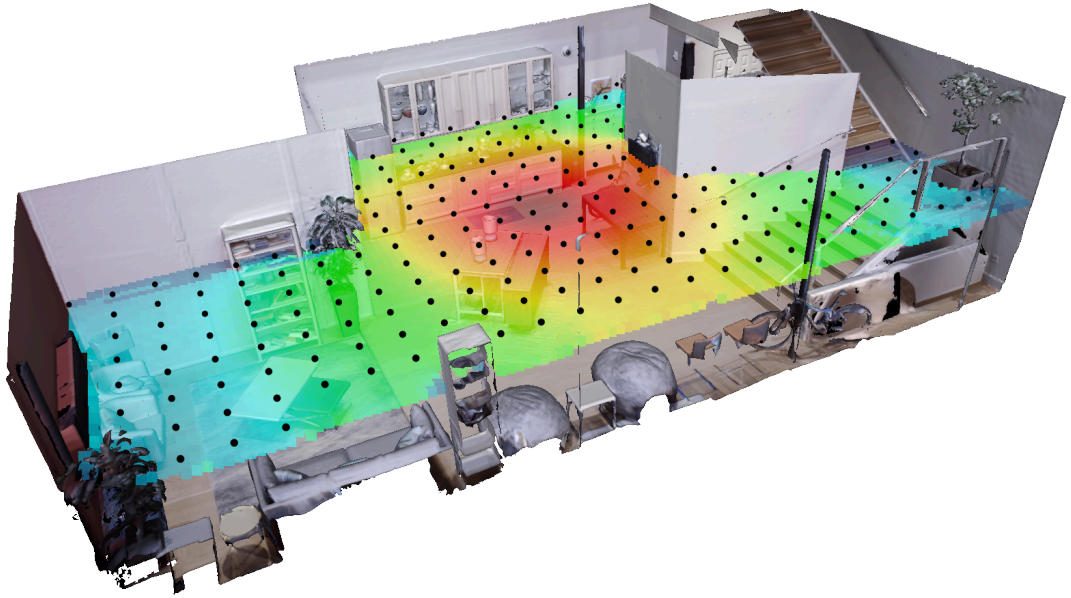
But today's embodied agent is **deaf**

We are the first to introduce audio-visual embodied navigation

C. Chen*, U. Jain*, et al., SoundSpaces: Audio-Visual Navigation in 3D Environments, ECCV 2020

# SoundSpaces: Our Audio Simulator

- We introduce SoundSpaces, an audio simulation platform to enable audio-visual navigation for two visually realistic 3D environments: Replica[1] and Matterport3D[2]



| | # Scenes | Avg. Area | # Training Eps. |
|---|---|---|---|
| Replica | 18 | 47.24 $m^2$ | 0.1M |
| Matterport3D | 85 | 517.34 $m^2$ | 2M |

Table: Summary of dataset statistics

[1] The Replica Dataset: A Digital Replica of Indoor Spaces, Straub et al., arXiv, 2019
[2] Matterport3D: Learning from RGB-D Data in Indoor Environments, Chang et al., 3DV, 2017

C. Chen*, U. Jain*, et al., SoundSpaces: Audio-Visual Navigation in 3D Environments, ECCV 2020

# SoundSpaces: Our Audio Simulator

- We introduce SoundSpaces, an audio simulation platform to enable audio-visual navigation for two visually realistic 3D environments: Replica[1] and Matterport3D[2]

- Our audio simulator produces realistic audio rendering based on the room geometry, materials, and sound source location

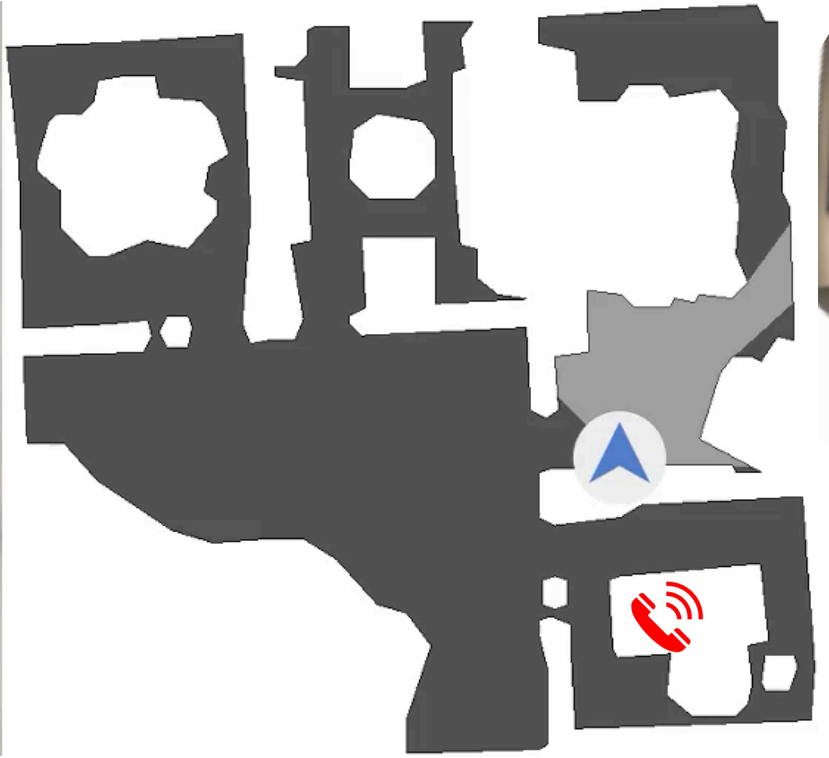
- The platform can play varying sounds of your choice in real time by precomputing a transfer function between locations

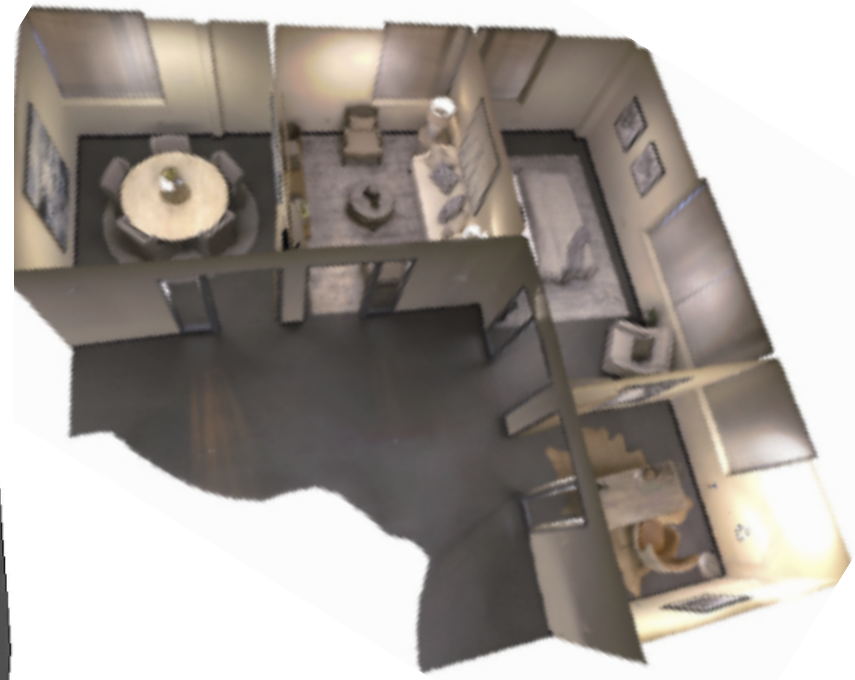
[1] The Replica Dataset: A Digital Replica of Indoor Spaces, Straub et al., arXiv, 2019
[2] Matterport3D: Learning from RGB-D Data in Indoor Environments, Chang et al., 3DV, 2017

# Where Is My Phone?

Agent view

Top-down map



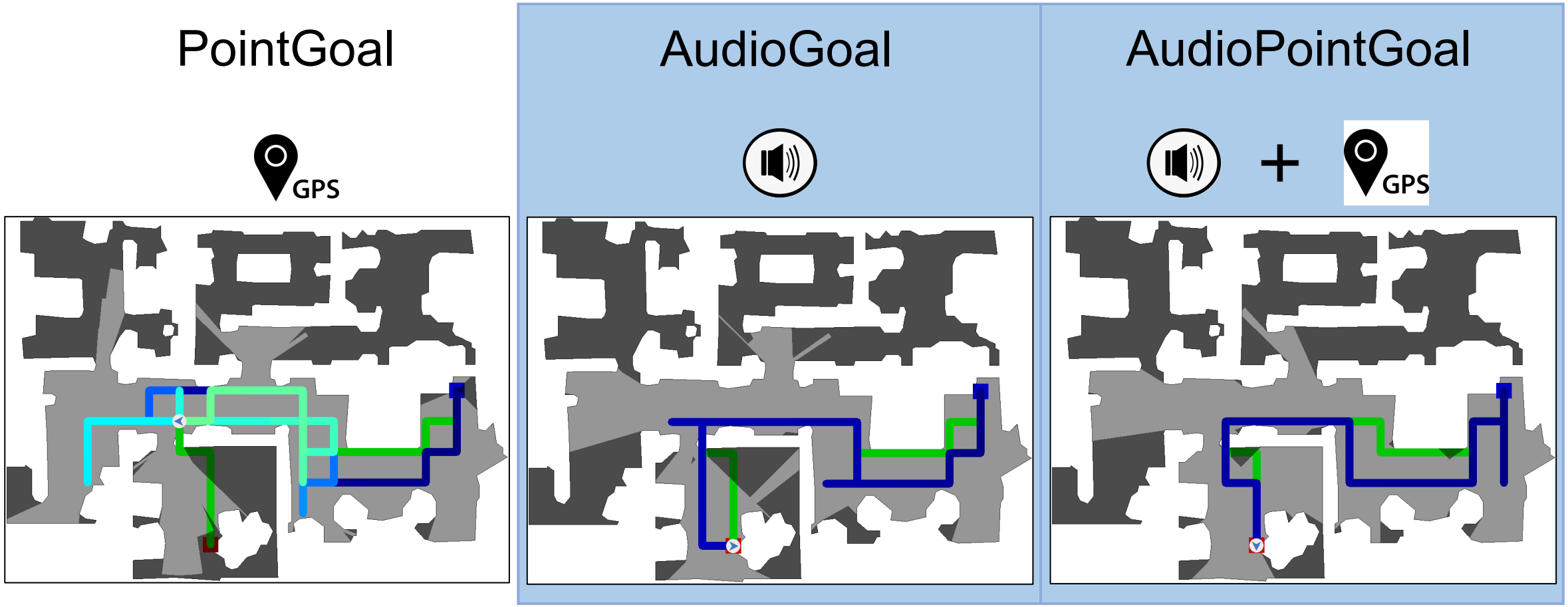


Direction: initially left ear is louder indicating sound coming from the left
Intensity: overall intensity gets higher as the agent gets closer to the goal

Agent | Goal | Start | Shortest path | Agent path | Seen/Unseen area | Occupied area

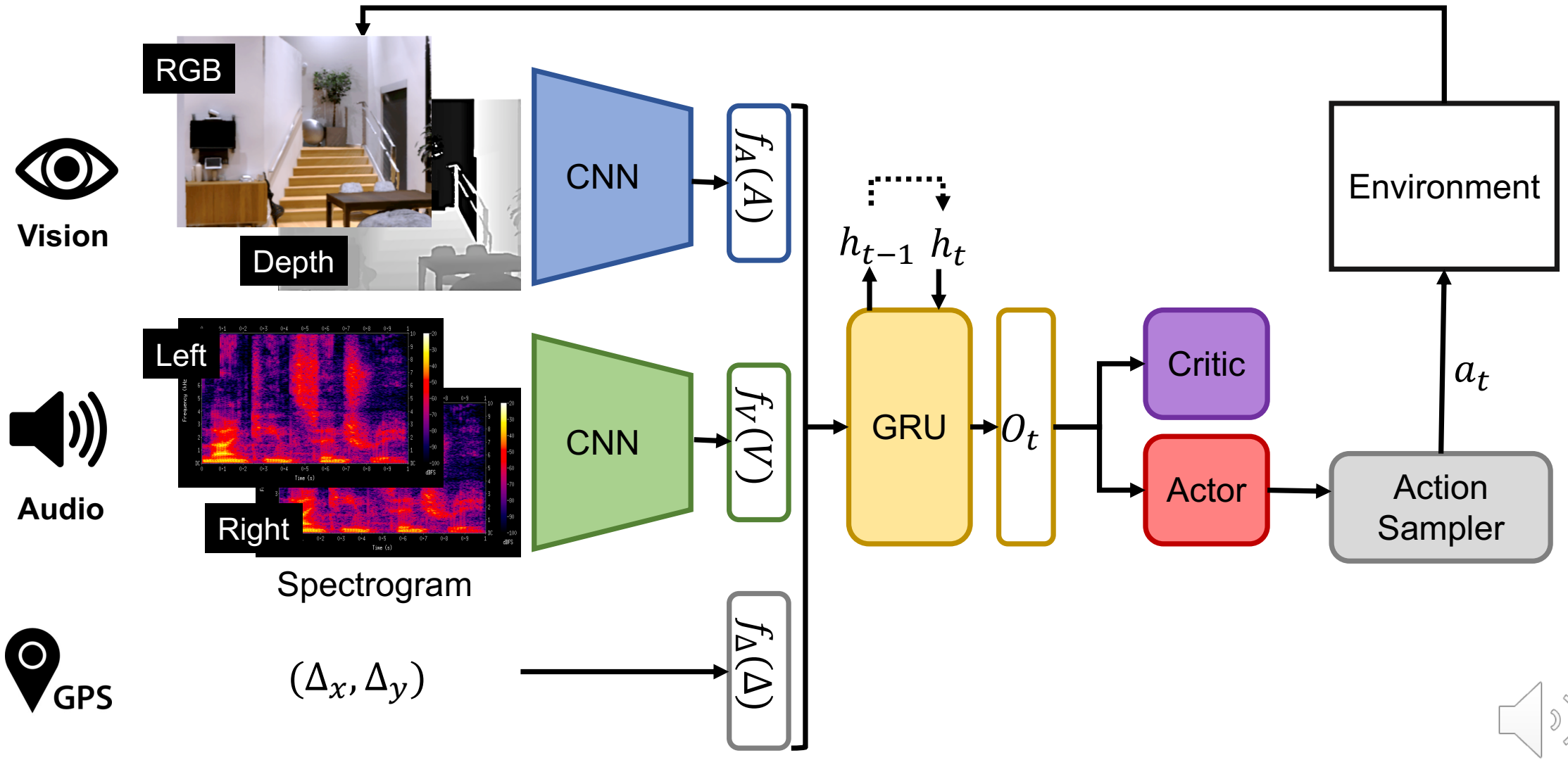
C. Chen*, U. Jain*, et al., SoundSpaces: Audio-Visual Navigation in 3D Environments, ECCV 2020

# Audio-Visual Navigation Tasks




C. Chen*, U. Jain*, et al., SoundSpaces: Audio-Visual Navigation in 3D Environments, ECCV 2020

# Deep RL for Audio-Visual Navigation




C. Chen*, U. Jain*, et al., SoundSpaces: Audio-Visual Navigation in 3D Environments, ECCV 2020

# Experiment Results

We show:
- Audio helps navigation
- Audio supplants GPS for and audio target
- Our agent generalizes to unheard sounds

Table 3: Navigation performance (SPL) when generalizing to unheard sounds. Higher is better. Results are averaged over 7 test runs; all standard deviations are $\leq 0.01$.

| Dataset | | *PG* | *Same sound* *AG* | *APG* | *Varied heard sounds* *AG* | *APG* | *Varied unheard sounds* *AG* | *APG* |
|---|---|---|---|---|---|---|---|---|
| Replica | Blind | 0.480 | 0.673 | 0.681 | 0.449 | 0.633 | 0.277 | 0.649 |
| | RGB | 0.521 | 0.626 | 0.632 | 0.624 | 0.606 | 0.339 | 0.562 |
| | Depth | 0.601 | 0.756 | 0.709 | 0.645 | 0.724 | 0.454 | 0.707 |
| Matterport3D | Blind | 0.426 | 0.438 | 0.473 | 0.352 | 0.500 | 0.278 | 0.497 |
| | RGB | 0.466 | 0.479 | 0.521 | 0.422 | 0.480 | 0.314 | 0.448 |
| | Depth | 0.541 | 0.552 | 0.581 | 0.448 | 0.570 | 0.338 | 0.538 |

# SoundSpaces: Audio-Visual Navigation in 3D Environments

*Changan Chen*[1,4], Unnat Jain*[2,4], Carl Schissler[3], Sebastia V. Amengual Gari[3], Ziad Al-Halah[1], Vamsi K. Ithapu[3], Philip Robinson[3], Kristen Grauman[1,4]*

*[1]UT Austin, [2]UIUC, [3]Facebook Reality Labs, [4]Facebook AI Research*

For more details, please refer to our paper:
https://arxiv.org/pdf/1912.11474.pdf

For more simulation demo, please check our project page:
http://vision.cs.utexas.edu/projects/audio_visual_navigation