Sounds Real: Using Hardware Accelerated Real-time Ray Tracing for Augmenting Location Dependent Audio Samples

Authors: Alexander Madzar, Linfeng Li, Hymalai Bello, Bo Zhou and Paul Lukowicz

Presenter: Bo Zhou
TU Kaiserslautern and DFKI
Kaiserslautern, Germany
Email: bo.zhou@dfki.de

IARIA

Resume of Dr. Zhou

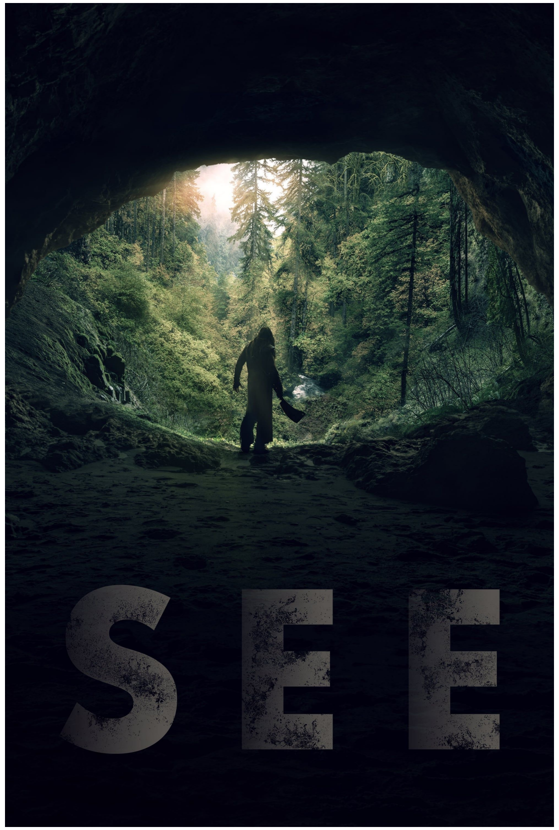
- Senior researcher in the German Research Center for Artificial Intelligence (DFKI)
- PhD in sensing and machine learning
- Recent research interest:
  - Transfer learning in sensor-based activity recognition

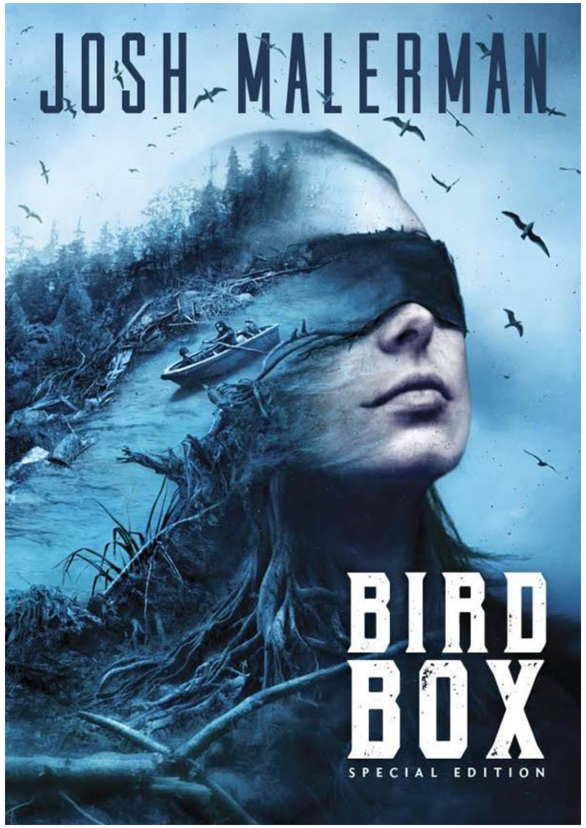
- https://www.drzhou.work/
- https://www.dfki.de/en/web/about-us/employee/person/bozh01/

Embedded Intelligence Research Group

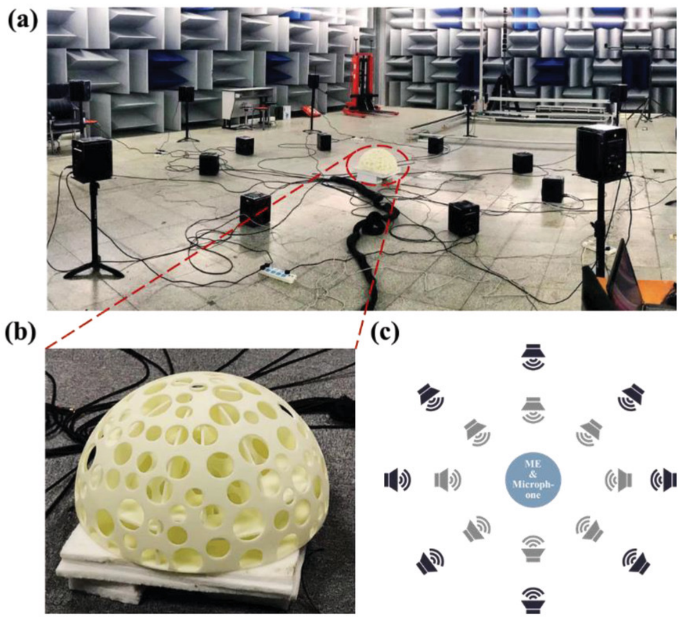
1. IoT, multi-sensor embedded systems
2. Wearable and pervasive sensing, activity recognition
3. Sensor fusion
4. Sensor data augmentation from 3D simulations

Sound is directional

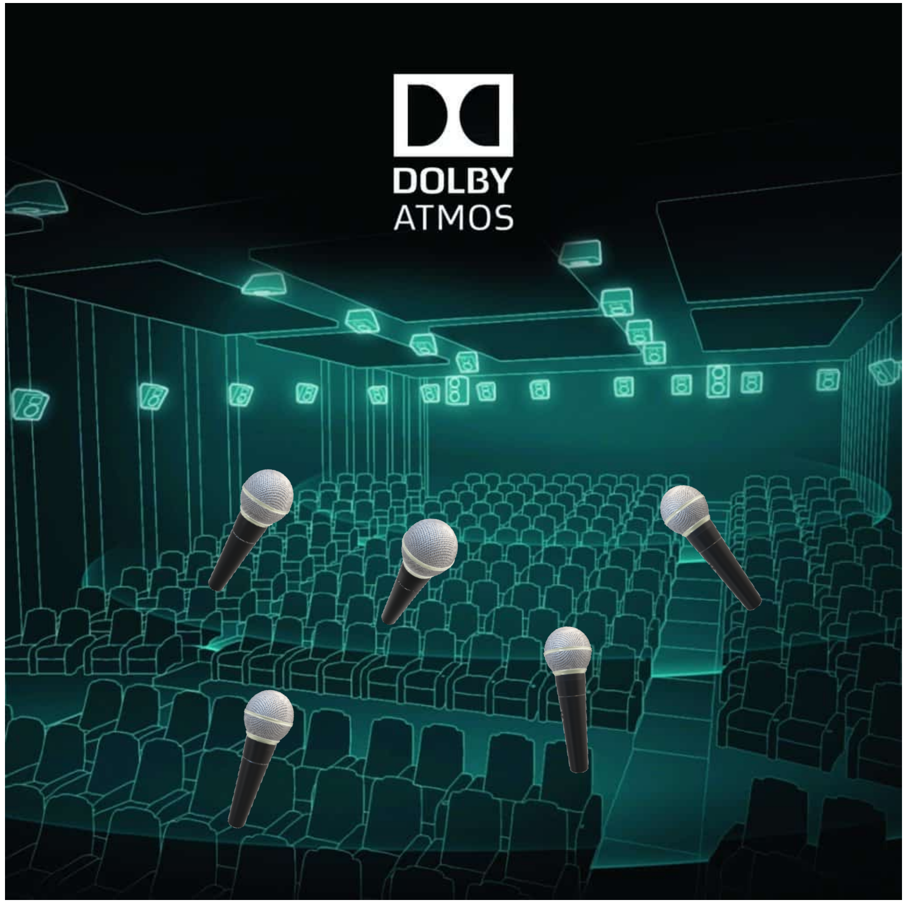




Sound Localization



Locate the source

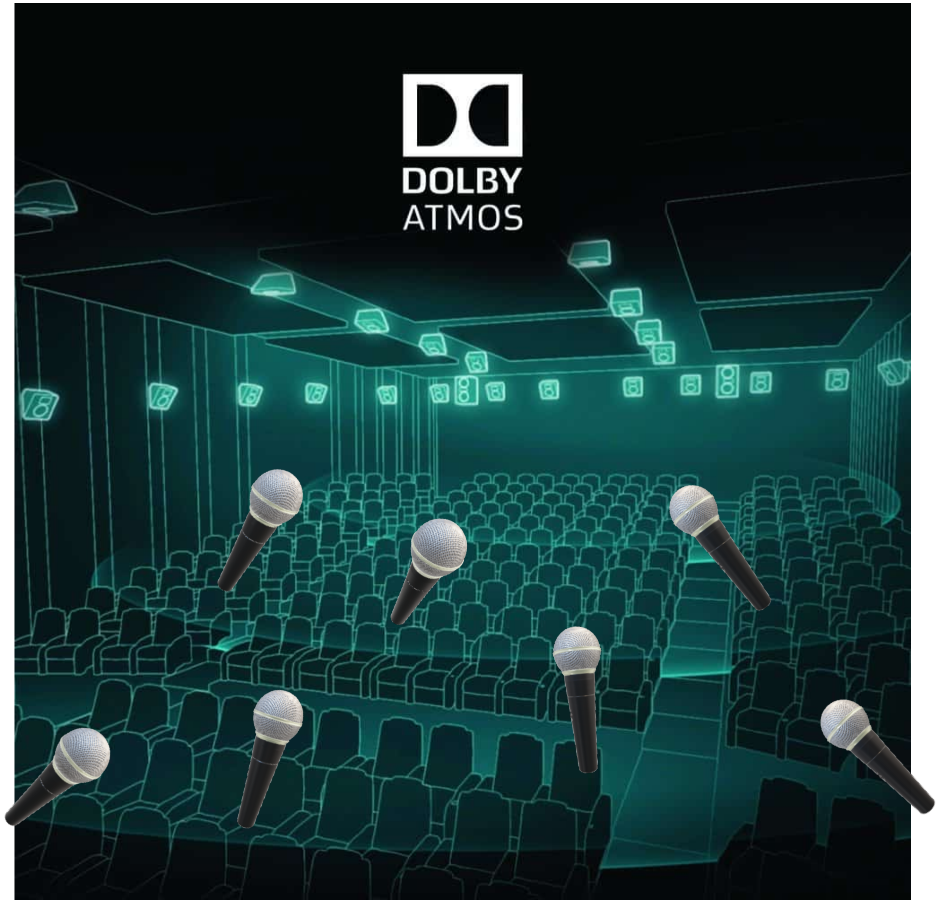


Locate the device?

Sun, Xuecong, et al. "Sound localization and separation in 3D space using a single microphone with a metamaterial enclosure." *Advanced Science* 7.3 (2020): 1902271.

Locating the listener

To train ML models to recognize the **listening device's location**, we will need experiment **data** with microphones at **every** point of interest in the field **at the same instance**



- This quickly becomes practically difficult
- And it's impossible to add more point of interests after the experiment

The 'location' can be definitive (as in coordinates) or abstract (e.g. close to the window)

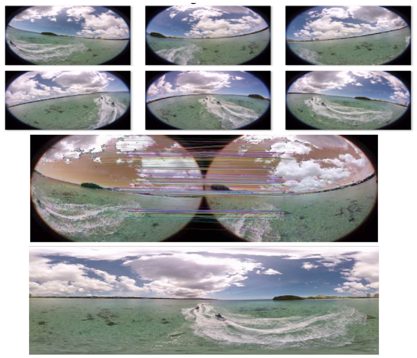
What if we can simulate the scene?

- What we need: a virtual recreation of a real-world scene
- We record the sound happening in the real world, and play back at their source location in the virtual scene
- Then we can insert virtual listeners at any location, any time
- We just need a simulation environment that has relatively precise sound engine, which already exists today

concept illustration from Cyberpunk 2077

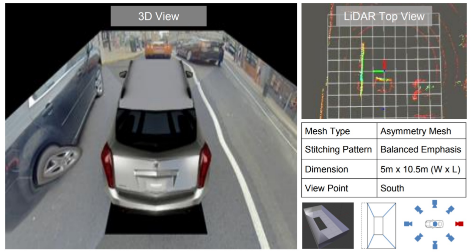
3D simulations in data augmentation

Automatic stitching is performed using Nvidia VRWorks 360 SDK.



O. Zia, J.-H. Kim, K. Han, and J. W. Lee, "360 panorama generation using drone mounted fisheye cameras," in 2019 IEEE International Conference on Consumer Electronics (ICCE), pp. 1–3, IEEE, 2019.

Algorithm for creating a 3D reconstruction of the Surround View in Cars for parking and unparking purposes.

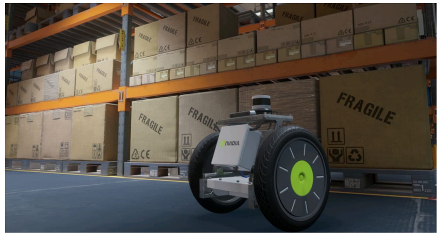


I. Baek, A. Kanda, T. C. Tai, A. Saxena, and R. Rajkumar, "Thinplate spline-based adaptive 3d surround view," in 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 586–593, IEEE, 2019

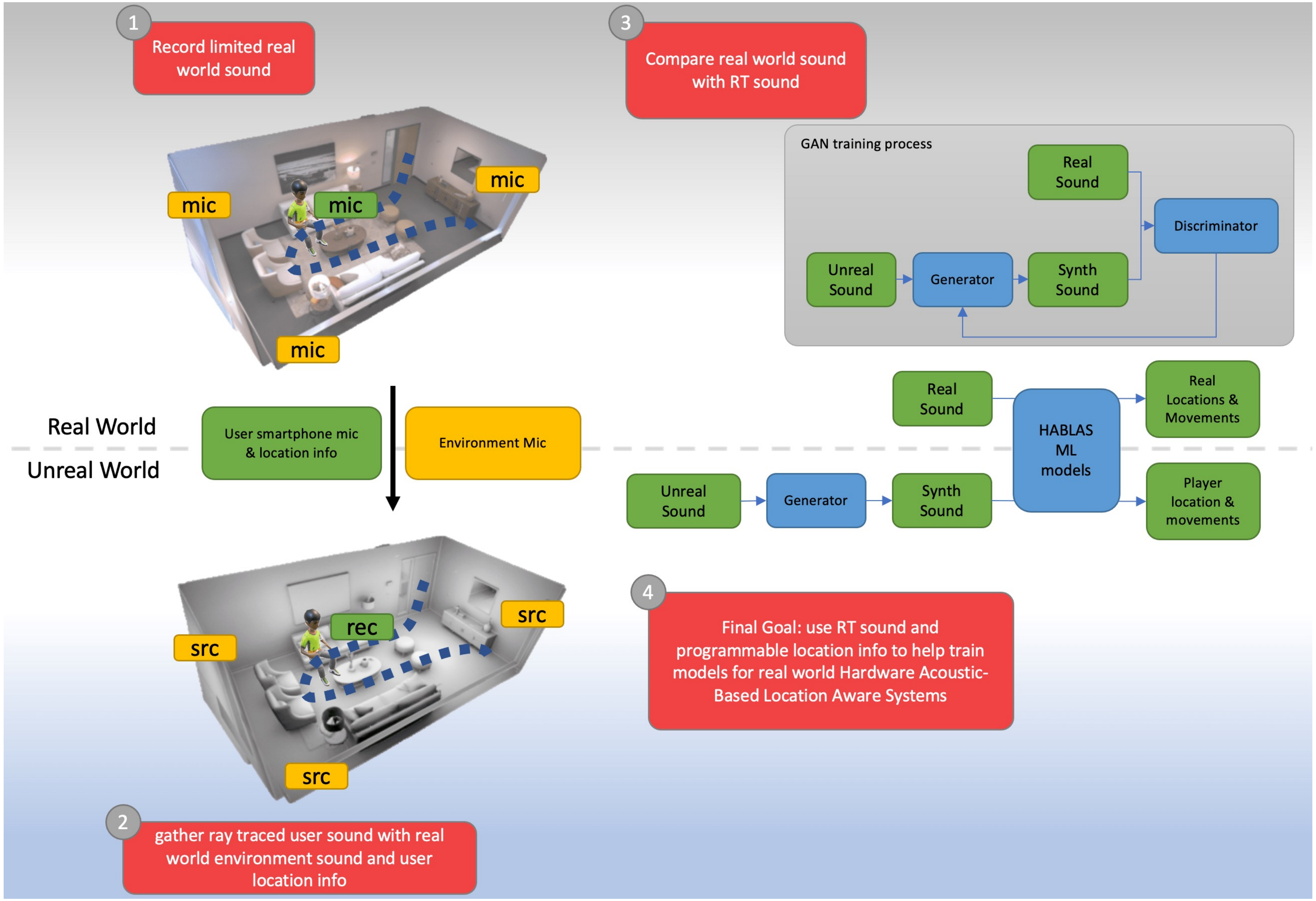
Photorealistic Virtual Reality Environment for Robotics Simulations and Synthetic Data Generation(in Unreal Engine 4)

P. Martinez-Gonzalez, S. Oprea, A. Garcia-Garcia, A. Jover-Alvarez, S. Orts-Escolano, and J. Garcia-Rodriguez, "**Unrealrox**: an extremely photorealistic virtual reality environment for robotics simulations and synthetic data generation," Virtual Reality, pp. 1–18, 2019.

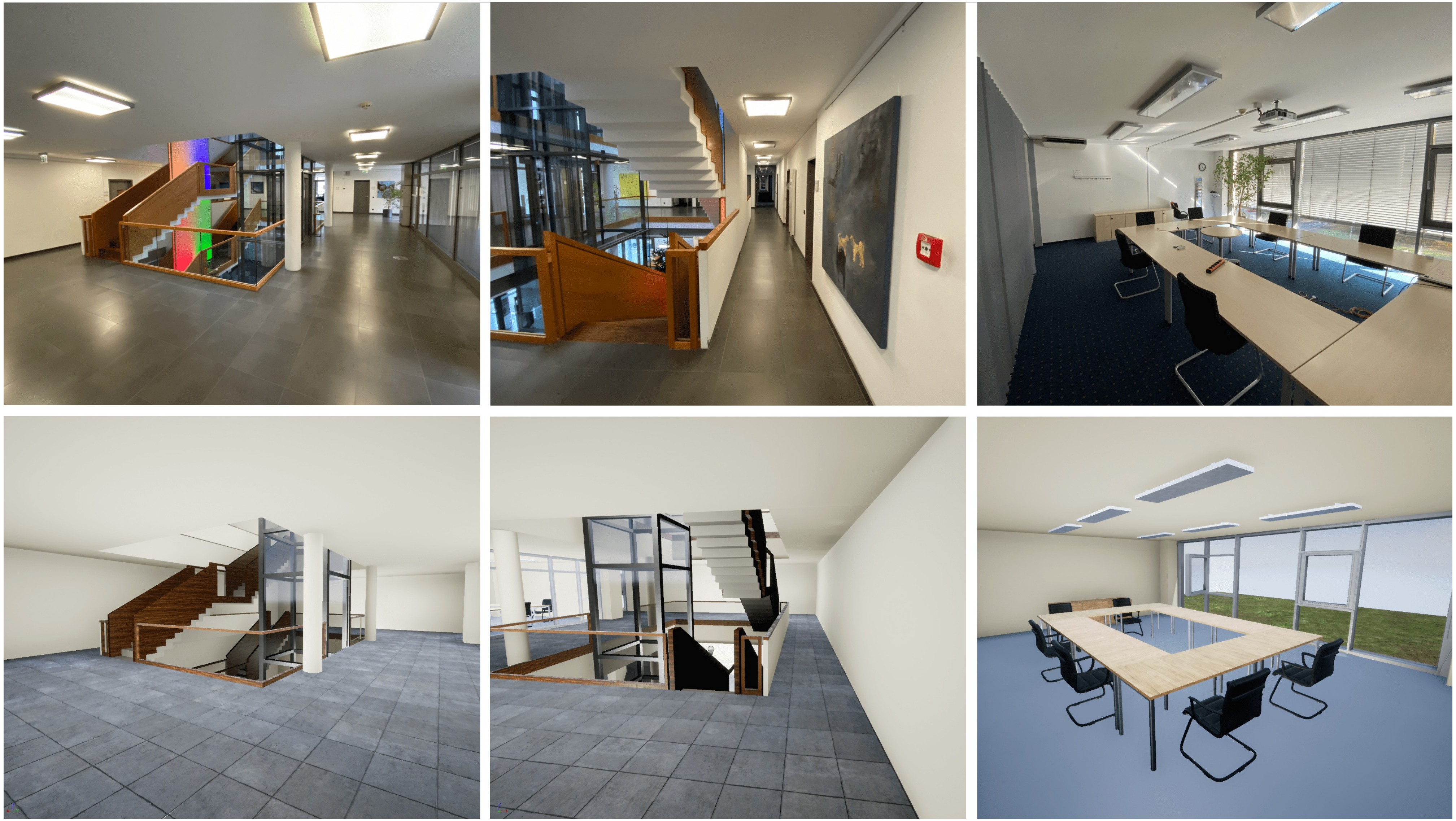
Revving Robotics: NVIDIA Isaac SDK Brings Modern AI to Autonomous Machines Robotics developer toolbox



https://blogs.nvidia.com/blog/2019/03/18/isaac-sdk-general-availability/


1
Record limited real world sound
mic
mic
mic
mic
3
Compare real world sound with RT sound
GAN training process
Real Sound
Discriminator
Unreal Sound
Generator
Synth Sound
Real World
Unreal World
User smartphone mic & location info
Environment Mic
Real Sound
HABLAS ML models
Real Locations & Movements
Unreal Sound
Generator
Synth Sound
Player location & movements
src
rec
src
src
4
Final Goal: use RT sound and programmable location info to help train models for real world Hardware Acoustic-Based Location Aware Systems
2
gather ray traced user sound with real world environment sound and user location info

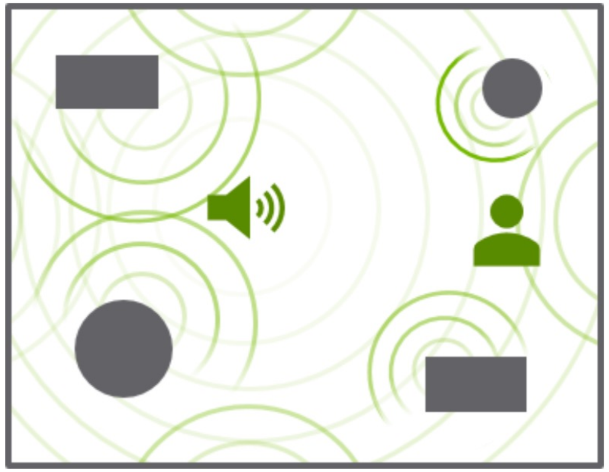
RECREATE SCENE



Framework

- Unreal Engine 4 (UE4) 4.15 is used to model the scene setup in combination with Nvidia VRWorks Audio.
- The sound sources were calibrated according to the following considerations:
  - Attenuate
  - Spatialize
  - Distance algorithm
  - Attenuation Shape
  - Radius
  - Falloff Distance
  - Occlusion
  - Direct and Indirect Path Gain
  - Effect strength.

Propagation



Path Traced Audio

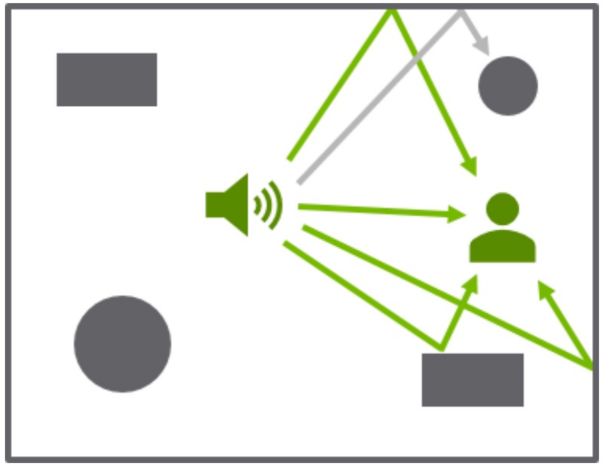
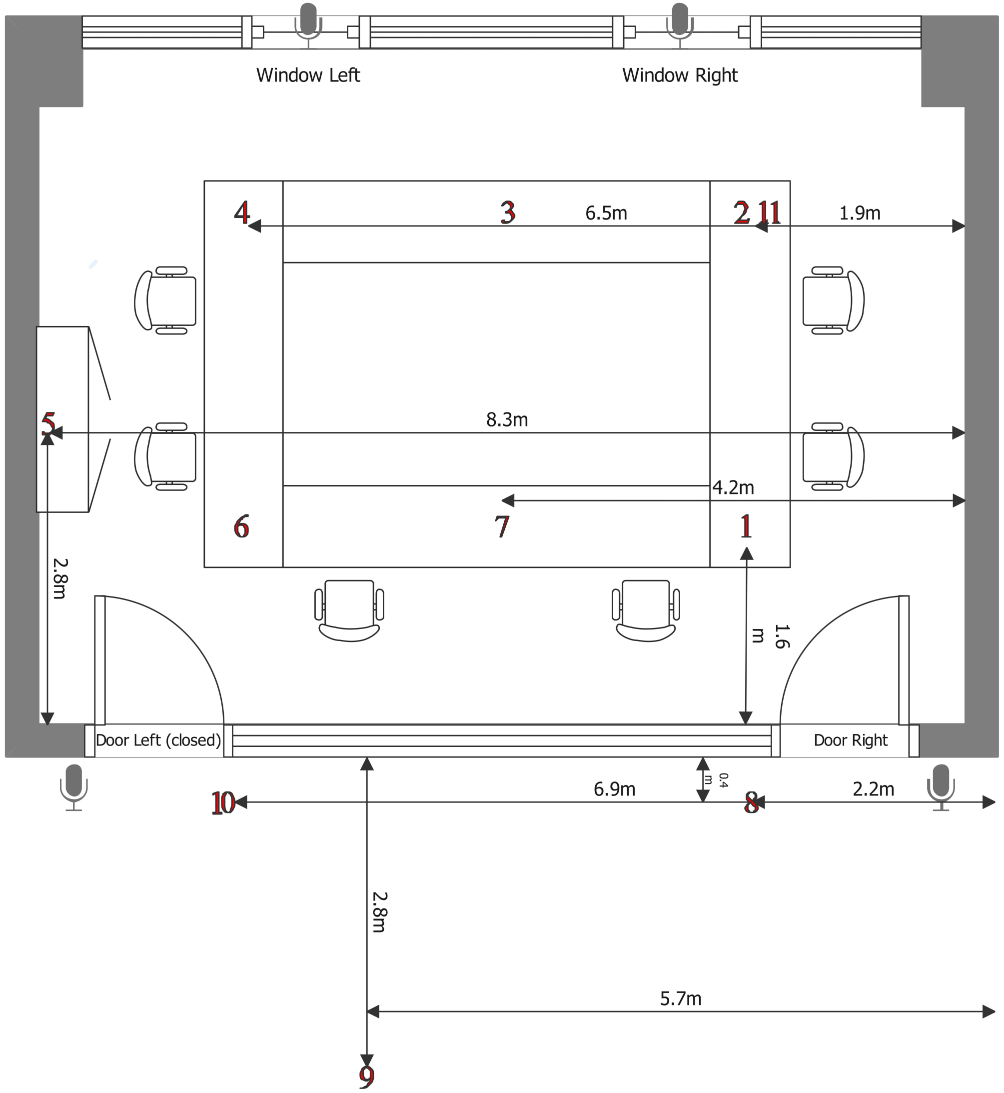


illustration from Nvidia VRWorks

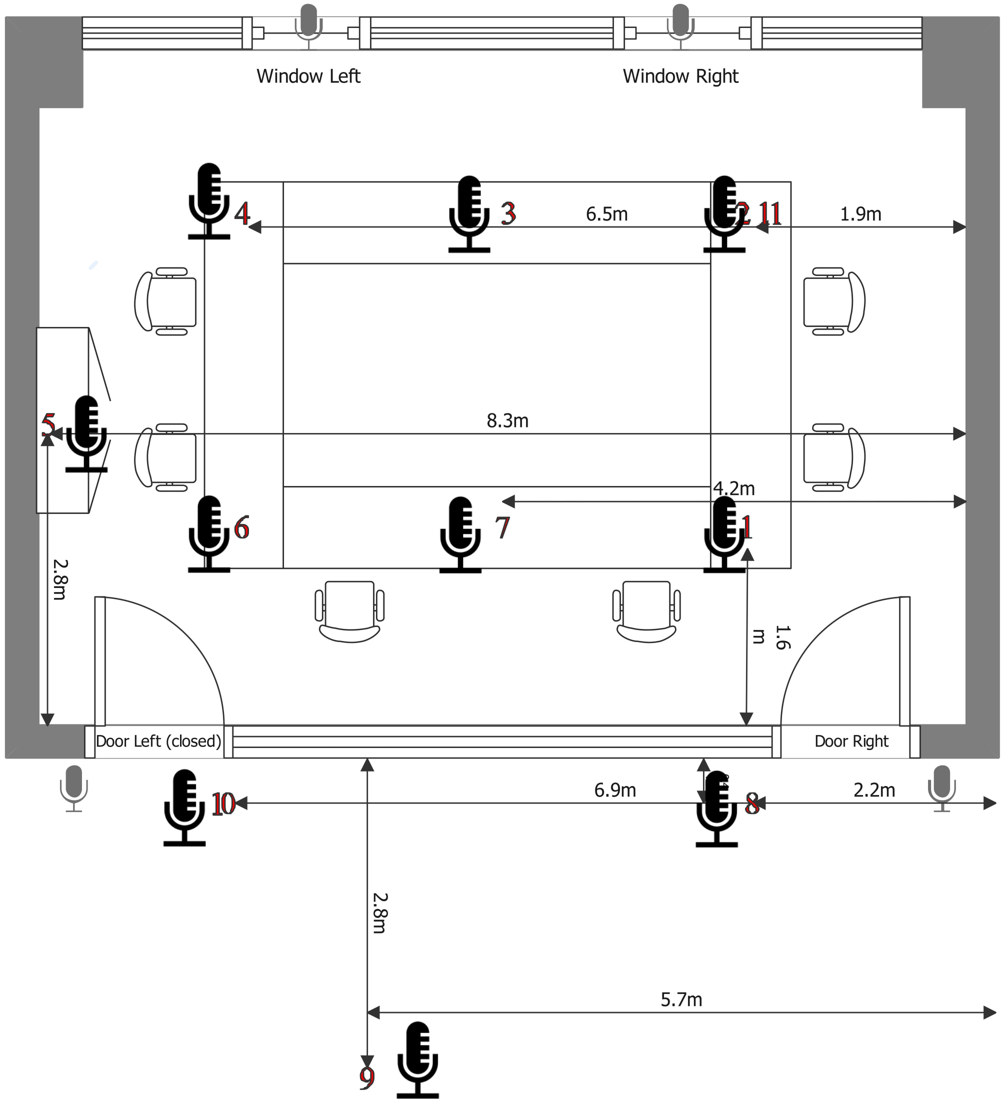
Methodology

A real-world experiment was designed to record ambient sound at predetermined places (1-11) and simulate the static positions of the avatar inside Unreal Engine.

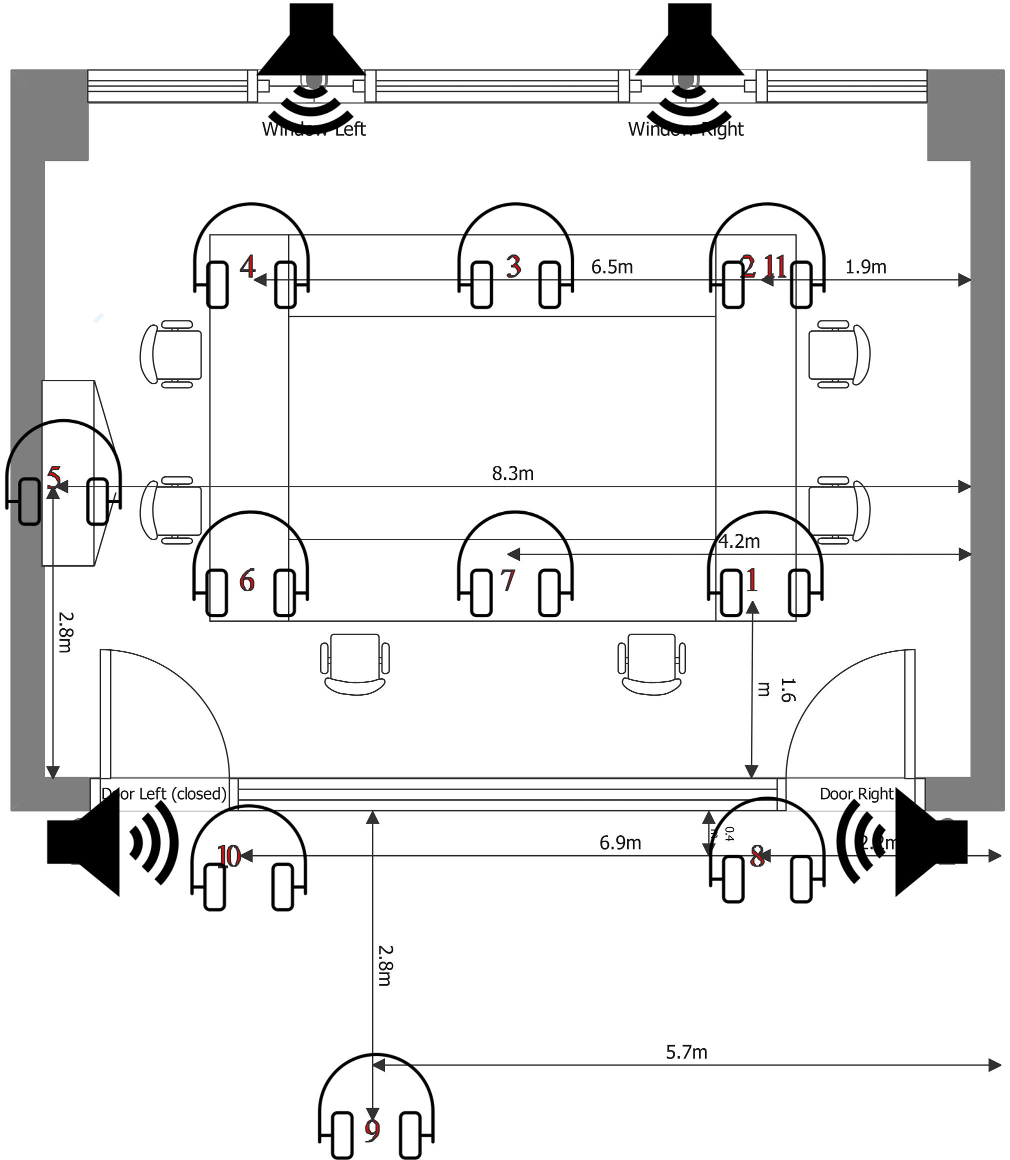
- The microphone symbols were used for gathering environment sound assets to play during real-time rendering.
- During the time, various activities were induced at random places in the scene, such as moving trolleys in the corridor, multiple people walking, powered drills, hammering metal, conversations, music, etc.
- The windows face a busy road, and there is always traffic sound captured at the two microphones at the windows.



data collection

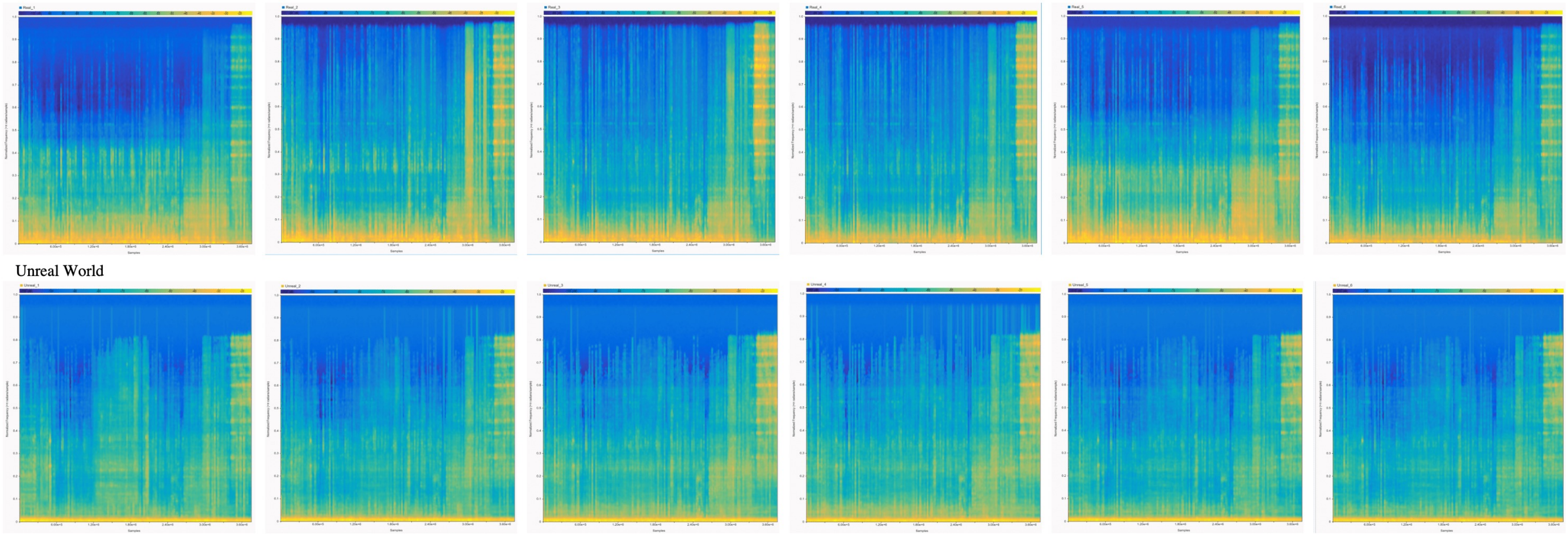


simulation



Real World

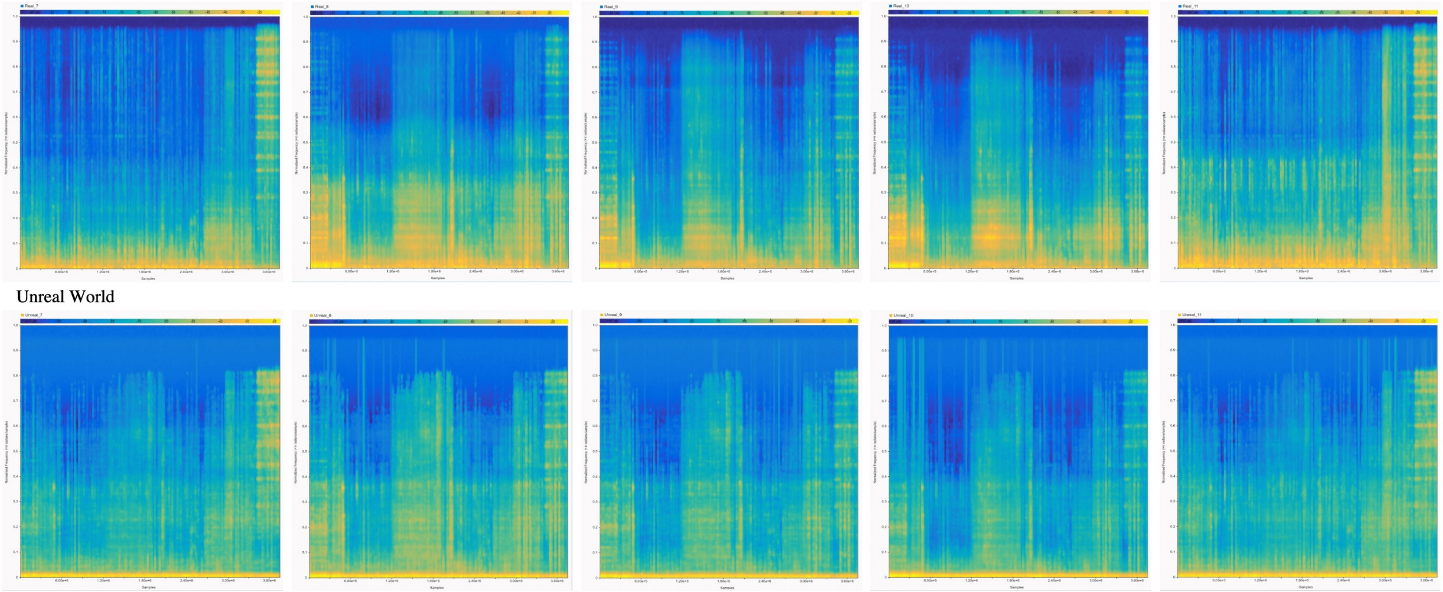
Unreal World



Position 1 Position 2 Position 3 Position 4 Position 5 Position 6

Real World

Unreal World



Position 7 Position 8 Position 9 Position 10 Position 11

Spectrogram of Real vs Unreal Worlds as Comparison

- Each position has a distinct appearance, although the sound played was the same in all 11 positions, as expected. It is indicating that the localization by the raytracing is valid.
- The virtual environment world’s spectrograms are less detailed in the high-frequency range compared to the real world.
  - Most likely due to the simulation environment, i.e., software restrictions, resolution of the digital sound card on the PC.
- Intense yellow colors can be noticed on the spectrograms of the virtual environment world. For instance, three events stand out at positions 8-10.
  - These are construction activities that occurred in the hallway adjacent to these positions.
  - However, these events are not discernible from the remaining locations, as construction work was barely audible within the meeting room.

Conclusion And Future Work

- Latest 3D game engines for data augmentation to generate acoustically realistic sound samples.
- This can be used further for training Hardware Audio Based Location-Aware Systems, bypassing difficulties in real-world data collection.
- Using our method, the scene can be replayed, and the listener can be placed at different locations for the same period of surrounding activities, which is not possible in real-world data collection.

- To be tested for location classification models
- Other frameworks like SteamVR.
- Post filtering with GAN
- Automatic scene creation like 3D cameras and semantic SLAM.

Thank you for your attention