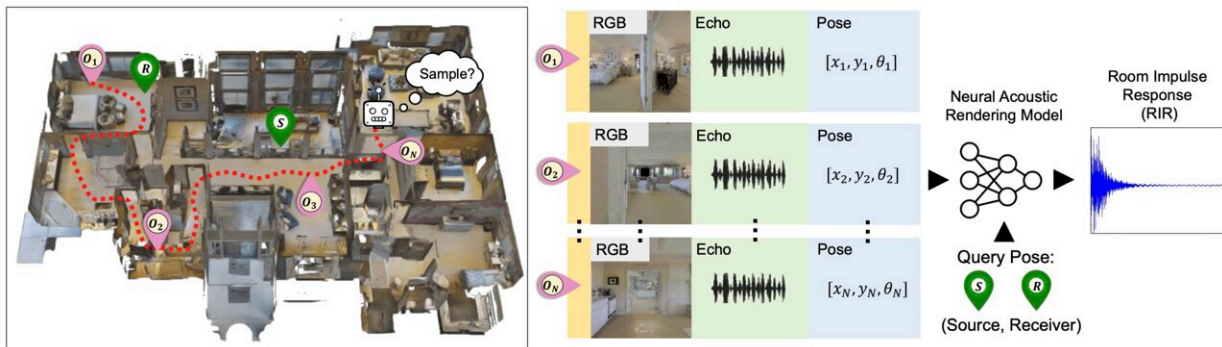# A new approach to efficiently model the acoustics of an environment

May 29 2024, by Ingrid Fadelli

Active acoustic sampling. An agent must intelligently navigate an unmapped 3D scene and actively sample audio-visual snapshots (the scene's acoustic context) to construct an acoustic model of the environment, given limited navigation time and a fixed sampling budget. When queried with an arbitrary sound source position and receiver location in the space, the learned environment acoustic model should accurately generate the corresponding Room Impulse Response (RIR) at that pose, where an RIR is a wave function that captures the acoustics for a particular scene, and a source and receiver location pair. Credit: Somayazulu et al.

Augmented reality (AR) and virtual reality (VR) are designed to artificially reproduce the experience of navigating certain environments. In recent years, videogame and entertainment companies have been producing increasingly immersive content that can be accessed using these technologies.

Some computational tools can aid in the creation of VR or AR content, allowing engineers to produce realistic models of real-world environments. These tools include so-called environment acoustic models, which are designed to reliably represent how sounds are transformed by the physical characteristics of different indoor environments.

Researchers at the University of Texas at Austin recently introduced ActiveRIR, a new approach to effectively estimate and model the acoustics of environments. This approach, introduced in a paper posted to the preprint server *arXiv*, utilizes reinforcement learning to produce high-quality acoustic models relying on only a few acoustic samples.

"We have been interested in the topic of efficient estimation of environment acoustics for some time now," Arjun Somayazulu and Sagnik Majumder, co-authors of the paper, told Tech Xplore.

"In this context, 'efficiency' refers to the notion of using a limited set of acoustic measurements in a novel 3D environment for estimating the acoustics of the entire scene. Estimating scene acoustics can facilitate AR/VR applications, where one wants to render spatially appropriate sounds for a 3D scene."

Conventional approaches for the modeling of acoustics can only make reliable estimations after analyzing a large quantity of audio samples collected from the environment of interest. This makes them impractical, as they would drain the battery of VR/AR devices and would require a long time to perform estimations.

"With this motivation in mind, we first proposed the idea of few-shot audio-visual learning of environment acoustics, where the goal is to predict the scene acoustics using a very few audio-visual samples from it," Somayazulu and Majumder explained.

2/5

"However, this and other concurrent work are limited in that it randomly picks a few points in scene for collecting the samples, which could be suboptimal as the randomly chosen points might not be the best set of the samples in terms of capturing the overall scene acoustics.

"Besides, they assume prior knowledge of the environment floorplan, which might not be available for previously unseen environments, and ignore the time and energy that it would take to physically cover all randomly chosen points, making it a bit disconnected from real-world applications."

As part of their recent study, Somayazulu and Majumder set out to address the limitations of their previously proposed method to model environmental acoustics, using a new task known as active acoustic sampling. This task entails the use of an embodied agent that moves around in an unknown 3D environment, while actively deciding where to collect audio-visual samples that would best aid the estimation of environment acoustics.

"The agent operates under both a time budget and a sample budget," Somayazulu and Majumder said. "While the time budget ensures that the agent navigates efficiently, the sample budget ensures that the agent does not collect samples that do not provide significant information about the environment acoustics. The combination of these two budgets improves the efficiency of the acoustics estimation task by limiting the time and energy used for the task."

The ActiveRIR model, the approach introduced by this research team, consists of two primary components that complement each other. The first is an audio-visual sampling policy, while the second is an acoustics estimation model.

"The sampling policy takes first-person audio-visual snapshots of the

environment and makes two important decisions: a) how to move in the scene, and b) where to collect a sample for estimating the scene acoustics," Somayazulu and Majumder said.

"The acoustics estimation model takes these samples and continuously improves its estimate of the overall scene acoustics. These two components share a symbiotic relationship."

The two components of ActiveRIR work closely together to ultimately produce realistic environment acoustic models. The sampling policy shares the most informative audio-visual samples with the acoustics estimator, allowing it to reliably estimate the acoustics of a given environment. In turn, the acoustics estimator helps the sampling policy to direct an embodied agent to places where collecting samples would be most beneficial to acoustic predictions.

The researchers evaluated their approach in a series of tests, comparing its performance to that of other techniques for estimating acoustics. They found that their sampling policy performed far better than many existing methods for motion planning and acoustics sample collection, including state-of-the-art techniques that learn to collect samples at new locations within a scene.

"Our framework is modular and generalizable enough to support multiple different acoustics estimation models, suggesting the possibility that it can be used to improve the sample-efficiency of any existing off-the-shelf model of your choice, while minimally compromising on its acoustics estimation quality," Somayazulu and Majumder said.

The new approach introduced by this team of researchers could soon be tested in a broader variety of settings using different agents to collect audio-visual samples. Eventually, it could contribute to the production of more VR and AR content that realistically reproduces the sounds of

specific 3D scenes.

"So far, we tested our model in a highly realistic indoor scene simulation platform," Somayazulu and Majumder added. "Looking ahead, however, it would be interesting to explore bridging the gap between simulation and the real world by evaluating how well ActiveRIR does on a physical robot in a real indoor space."

More information: Arjun Somayazulu et al, ActiveRIR: Active Audio-Visual Exploration for Acoustic Environment Modeling, *arXiv* (2024). DOI: 10.48550/arxiv.2404.16216

Citation: A new approach to efficiently model the acoustics of an environment (2024, May 29) retrieved 20 June 2024 from https://techxplore.com/news/2024-05-approach-efficiently-acoustics-environment.html