Measuring Acoustics with Collaborative Multiple Agents*

Yinfeng Yu[1,6], **Changan Chen**[2], **Lele Cao**[1,3], **Fangkai Yang**[4] and **Fuchun Sun**[1,5,†]
[1]Beijing National Research Center for Information Science and Technology, State Key Lab on Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University
[2]University of Texas at Austin
[3]Motherbrain, EQT Group
[4]Microsoft Research
[5]THU-Bosch JCML Center
[6]College of Information Science and Engineering, Xinjiang University
yyf17@mails.tsinghua.edu.cn, changan@cs.utexas.edu, lele.cao@eqtpartners.com,
fangkai.yang@microsoft.com, fcsun@mail.tsinghua.edu.cn

Abstract

As humans, we hear sound every second of our life. The sound we hear is often affected by the acoustics of the environment surrounding us. For example, a spacious hall leads to more reverberation. Room Impulse Responses (RIR) are commonly used to characterize environment acoustics as a function of the scene geometry, materials, and source/receiver locations. Traditionally, RIRs are measured by setting up a loudspeaker and microphone in the environment for all source/receiver locations, which is time-consuming and inefficient. We propose to let two robots measure the environment's acoustics by actively moving and emitting/receiving sweep signals. We also devise a collaborative multi-agent policy where these two robots are trained to explore the environment's acoustics while being rewarded for wide exploration and accurate prediction. We show that the robots learn to collaborate and move to explore environment acoustics while minimizing the prediction error. To the best of our knowledge, we present the very first problem formulation and solution to the task of collaborative environment acoustics measurements with multiple agents.

1 Introduction

Sound is critical for humans to perceive and interact with the environment. Before reaching our ears, sound travels via different physical transformations in space, such as reflection, transmission and diffraction. These transformations are characterized and measured by a Room Impulse Response (RIR) function [Välimäki *et al.*, 2016]. RIR is the transfer function between the sound source and the listener (microphone). Convolving the anechoic sound with RIR will get the sound with reverberation [Cao *et al.*, 2016]. RIR is utilized in many applications such as sound rendering [Schissler *et al.*, 2014], sound source localization [Tang *et al.*, 2020], audio-visual matching [Chen *et al.*, 2022], and audio-visual navigation [Chen *et al.*, 2020; Chen *et al.*, 2021b; Chen *et al.*, 2021a; Yu *et al.*, 2022b]. For example, to achieve clear speech in a concert hall, one might call for a sound rendering that drives more acoustic reverberation while keeping auditoriums with fewer reverberation [Mildenhall *et al.*, 2022]. The key is to measure RIR at different locations in the hall. However, RIR measuring is time-consuming due to the large number of samples to traverse. To illustrate, in a $5{\times}5$ m^2 room with a spatial resolution of 0.5m, the number of measurable points is $11{\times}11{=}121$. The source location (omnidirectional) can sample one of these 121 points. Assuming a listener with four orientations (0, 90, 180, 270), this listener can choose from 121 points with four directions for each chosen point. So, the number of source-listener pairs becomes $121{\times}121{\times}4{=}58{,}564$. Assuming the sampling rate, duration and precision of binaural RIR is 16K, 1 second and float32 respectively, one RIR sample requires $2 \times 16000 \times 4$ Bytes = 128KB from computer storage (memory). The entire room would take up to $58,564 \times 128$ KB ≈ 7.5 GB. Moreover, it also means that one has to move the source/listener devices 58,564 times and performs data sending/receiving for each point.

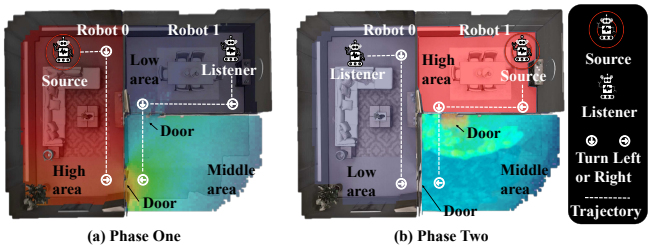

Figure 1: Learn to measure environment acoustics with two collaborative robots. The background color indicates sound intensity ("High", "Middle" and "Low" areas). Each step (one step per second) embodies three steps: 1) robot 0 emits a sound, and robot 1 receives the sound; 2) robot 1 emits the sound, and robot 0 receives the sound; 3) two robots make a movement following their learned policies. This process repeats until reaching the maximum number of time steps.

*The full paper with appendix together with source code can be found at https://yyf17.github.io/MACMA.

†Corresponding author: Fuchun Sun.