Deep Neural Room Acoustics Primitive

Yuhang He[1], Anoop Cherian[2], Gordon Wichern[2], Andrew Markham[1]
[1]Department of Computer Science, University of Oxford, UK
[2]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, US

1. Problem Definition

Goal: Learn a spatially continuous neural RIR room acoustics field, so as to be able to

1. predict spatial acoustic effects from,
2. arbitrary source position to arbitrary receiver position.

by

sending and receiving sound at discrete positions.

2. Challenges

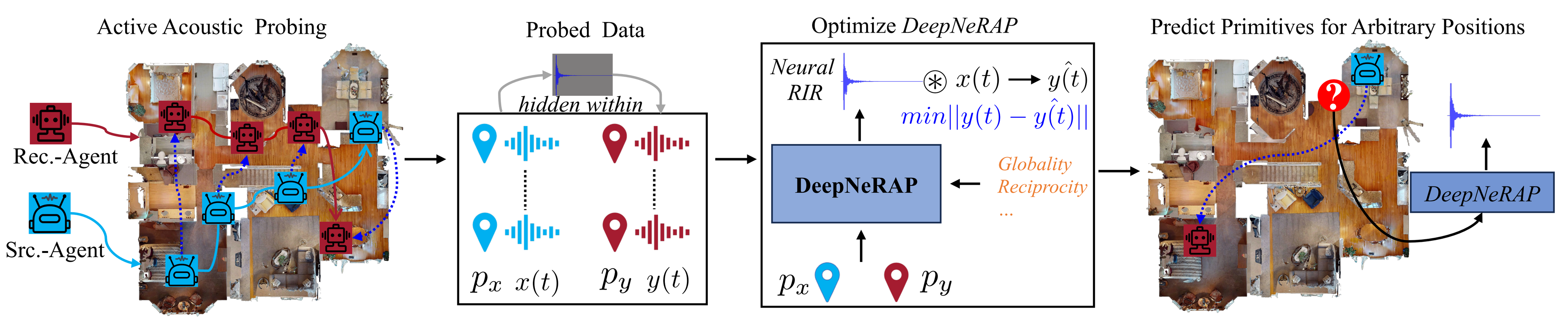
Room impulse response (RIR) is

1. highly non-smooth and chaotic, lengthy in points.
2. position-sensitive, smaller position change leads to large RIR change.

Measuring RIR in real-world,

1. is difficult and inefficient (time-consuming).
2. nonscalable, need re-collect once position changes.

3. Our Method: DeepNeRAP

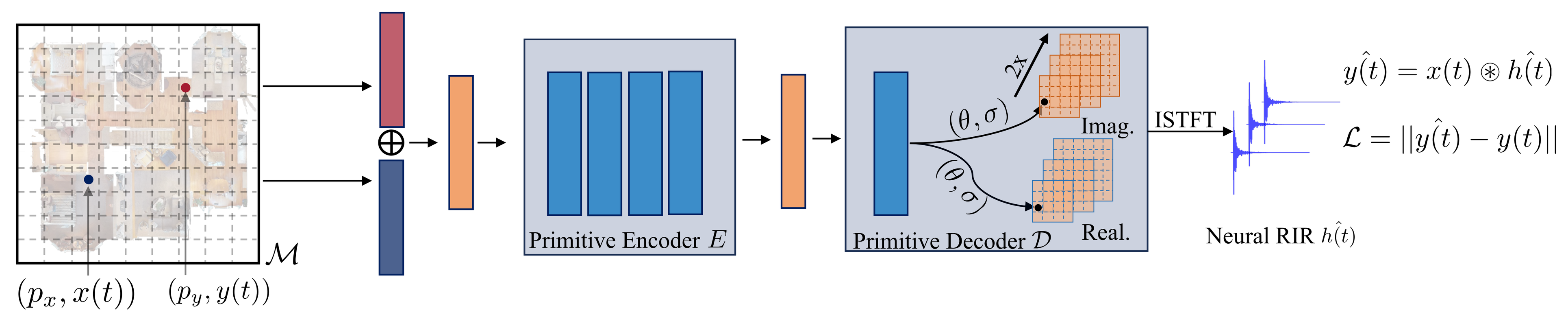


1. one source-agent and one receiver-agent.
2. walk around the room independently, sending and receiving sine-sweep sound at different positions.
3. learn neural RIR in a self-supervised manner.
4. neural RIR is optimized by minimizing discrepancy between neural RIR effected receiver sound and recorded receiver sound.
5. able to predict neural RIR for new arbitrary source-receiver positions.

4. Room Acoustics Physical Principles

1. **Globality**: sound propagation relates to whole room.
2. **Reciprocity**: RIR is the same if source/receiver swaps.
3. **Superposition**: multiple sound sources equals to adding each one together.
4. **Sound Independence**: learned neural RIR field is independent on sound source class and existence.

5. Algorithm Pipeline



1. Extract multi-scale position aware feature (*globality*).
2. Element-wise add two features (*reciprocity*).
3. Predict neural RIR in frequency domain (real/imag).

6. Experiment Result

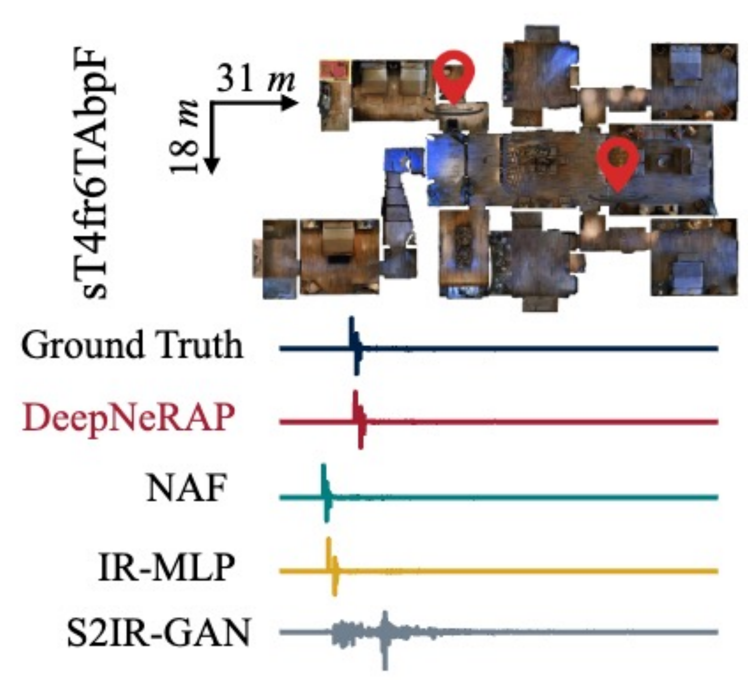
1. **Dataset**: Synthetic: SoundSpaces 2.0 [1] + MP3D [2], Real: MeshRIR data [3].

2. Quantitative Result (MP3D):

Table 1. Quantitative Result on Matterport3D Dataset. t-MSE: 10^{-7}, f-MSE: 10^{-2}.

Method	Neural RIR						Speech
	t-MSE (↓)	SDR (↑)	T_{60} Error (↓)	f-MSE (↓)	PSNR (↑)	SSIM (↑)	PSEQ (↑)
NAF (Luo et al., 2022)	1.01± 0.27	5.16± 0.09	7.84± 0.40	4.09 ± 0.00	15.17± 3.57	0.996± 0.00	1.40 ± 0.41
IR-MLP (Richard et al., 2022)	1.02± 0.32	4.09± 0.10	8.68± 0.12	5.68± 0.01	13.67± 3.51	0.994± 0.01	1.40 ± 0.14
S2IR-GAN (Ratnarajah et al., 2023)	1.09± 0.27	3.81± 0.10	9.19± 0.02	6.55 ± 0.12	12.98± 3.46	0.994± 0.01	1.38 ± 0.17
DeepNeRAP	**0.93 ± 0.34**	**6.62± 0.12**	**6.04± 0.08**	**1.68± 0.02**	**18.95± 3.02**	**0.998± 0.00**	**1.53 ± 0.41**

3. Qualitative Result:



Conclusion:

1. Novel and simple neural RIR learning framework.
2. Physical principles (interpretable result) informed network design.
3. SOTA performance.

[1] Changan Chen et al., SoundSpaces 2.0: A Simulation Platform for Visual-Acoustic Learning. NeurIPS 2022.
[2] Angel Chang et al, Matterport3D: Learning from RGB-D Data in Indoor Environments. 3DV 2017.
[3] Shoichi Koyama et al., MeshRIR: A Dataset of Room Impulse Responses on Meshed Grid Points For Evaluating Sound Field Analysis and Synthesis Methods. WASPAA. 2021.