COMP4801 Final Year Project

Detailed Project Plan

Title: Geometry Aware Room Impulse Response Generation and Visualization

Yu, Yuemin, UID: 3035945766
Department of Computer Science, HKU
Supervisor: Prof. Wu, Chenshu

1 Project Background

When sound signals travel within an indoor environment, absorption, reflection, diffraction, and attenuation might occur due to interaction with the walls, floor, ceiling, etc. This phenomenon is represented by Room Impulse Response (RIR), which is a linear and causal time-domain filter[2]. RIR is a fundamental concept in acoustics and signal processing. It is widely used in different topics. For instance, when convolving clean speech with RIRs and adding background noise, a far-field speech training dataset can be synthesized[9].

1.1 Room Impulse Response Generation

Measuring RIR often requires significant computational resources and lengthy processing times, making it challenging to deploy on edge devices with limited capabilities. Therefore, RIR generation is an important task. There are several methods for generating RIRs currently. One of the methods is using image-source method. It was first proposed and developed in 1979[1], and Diaz-Guerra et al.[3] demonstrated an GPU-based accelerated library for this method. Another popular method is utilizing the power of machine learning to do this task, which will be discussed in Section 1.4.

1.2 Diffusion Models

Diffusion models are a class of generative models that have shown promising results in generative tasks. There are several famous model structures, for instance, Denoising Diffusion Probabilistic Models [5] and Generative modeling by estimating gradients of the data distribution[8]. Currently, most of the tasks conducted by such diffusion models are vision-based, including image inpainting, image generation (see Fig. 1), etc. This project aims to explore the potential of diffusion models in the audio domain.

1.3 Frontend Framework

There are quite a few frontend frameworks available for web development in recent years. React.js is a popular JavaScript library for building user interfaces. It is widely used in web

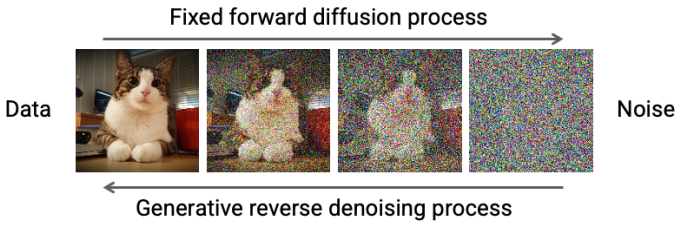

Figure 1: An example figure of image generation using diffusion models.

development and is known for its simplicity and flexibility. According to public data from StackOverflow[6] and some surveys, React.js is the most popular front-end framework in recent years. Many famous websites, such as Facebook, Instagram, and Netflix (see Fig. 2), are built using React.

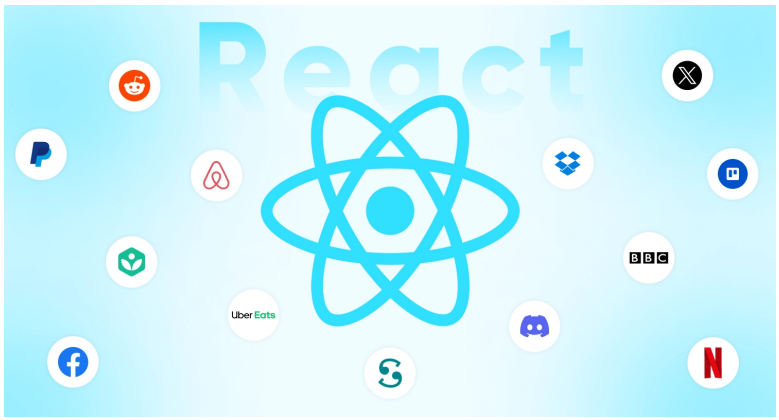

Figure 2: An example figure of image generation using diffusion models.

1.4 Related Work

There are already some works on RIR generation using the machine learning techniques. One of the most famous works is the Conditional Generative Adversarial Networks (cGAN) based method[7], which used cGAN as the foundation model, with significant improvements in the quality of the generated RIRs as well as the runtime. Another work is the implementation of diffusion-based model[4], however the author claimed that the performance of the model is not as good as the cGAN-based method. Hence this project will try to dive into the diffusion-based model and try to improve the performance.

2 Project Objective

The project aims to develop a novel method for room impulse response generation and visualize the output, based on web application. The main objectives are following:

- **A novel diffusion model:** The project will develop a diffusion-based deep learning model for RIR generation task.
- **Output based on custom input:** The project will generate outputs based on input parameters collected and provide the output in short runtime.
- **High-quality RIR generation:** The project will generate high-quality RIRs with high accuracy in various evaluation metrics, such as Reverberation Time Error and Word Error Rate when used in speech recoginition tasks.
- **A end-to-end system with high-quality visualization:** The project will aim to incorporate several efficient design strategies to ensure our algorithms can run effectively on ubiquitous edge devices. Meanwhile, visualization will be provided based on a Web application.

3 Project Methodology

3.1 Literature Review

The project will start with a comprehensive literature review on the related topics, including various fields from acoustics, signal processing, machine learning, to web development. State-of-the-art methods will be learned, and key findings will be summarized.

3.2 Dataset Collection

Previously used public datasets will be collected and preprocessed. Moreover, more data samples might be generated using previous off-the-shelf methods to increase the size of the dataset.

3.3 Model Training and Efficiency Improvement

- Model Training: The project will aim to utilize the new architectures of diffusion model of state of the art to do training. Hyperparameters will be validated and tuned based the model performance, custom loss functions will be designed, and Pytorch framework will be used for implementation.
- Efficiency Improvement: there might be modification on the structure of the network to improve efficiency.

As mentioned in Section 2, a number of evaluation metrics will be compared with the state-of-the-art methods, including but not limited to [4], [7] and [3].

3.4 System Development

- System Deployment: the developed model will be deployed and tested on ubiquitous devices, such as personal laptop or even Raspberry Pi and Jetson Nano, and then undergo specific design and tuning to ensure the efficiency of the model.
- Web application development: the project will use React.js as Frontend, and utilized popular Backend frameworks, such as Flask, Django, FastAPI or Spring Boot.

4 Project Schedule and Milestones

The tentative schedule for the project is as follows:

Timeline	Tasks
2024.09 - 2024.10	Literature review on related topics of the project, prepare the dataset with large enough size for training, validation and testing.
2024.11 - 2024.12	Develop the model, test and deploy the system on ubiquitous devices, seek user feedback.
2025.01 - 2025.02	Tune the model and system to get optimal output, develop the web application.
2025.03	Test the system and model, and evaluate the performance of the project.
2025.04 - 2025.05	Update the project webpage accordingly, finalize the report and prepare for the final presentation with all deliverables ready.

Table 1: Project Schedule and Milestones

References

[1] Jont B. Allen and David A. Berkley. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950, April 1979.

[2] Diego Di Carlo, Pinchas Tandeitnik, Cedrić Foy, Nancy Bertin, Antoine Deleforge, and Sharon Gannot. dechorate: a calibrated room impulse response dataset for echo-aware signal processing. *EURASIP Journal on Audio, Speech, and Music Processing*, 2021(1):1–15, 2021.

[3] David Diaz-Guerra, Antonio Miguel, and Jose R. Beltran. gpurir: A python library for room impulse response simulation with gpu acceleration. *Multimedia Tools and Applications*, 80(4):5653–5671, October 2020.

[4] Eric Grinstein and Zehua Chen. RoomFuser: Room Impulse Response Generation using Neural Diffusion Models, October 2023.

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *arXiv preprint arXiv:2006.11239*, 2020. Accessed: 2024-09-23.

[6] Stack Overflow. Tag trends of frontend frameworks. `https://trends.stackoverflow.co/?tags=reactjs,vue.js,angular,svelte,angularjs,vuejs3`, 2024. Accessed: 2024-09-23.

[7] Anton Ratnarajah, Shi-Xiong Zhang, Meng Yu, Zhenyu Tang, Dinesh Manocha, and Dong Yu. Fast-rir: Fast neural diffuse room impulse response generator. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 571–575, 2022.

[8] Yang Song and Stefano Ermon. *Generative modeling by estimating gradients of the data distribution*. Curran Associates Inc., Red Hook, NY, USA, 2019.

[9] Igor Szöke, Miroslav Skácel, Ladislav Mošner, Jakub Paliesek, and Jan Černocký. Building and evaluation of a real room impulse response dataset. *IEEE Journal of Selected Topics in Signal Processing*, 13(4):863–876, 2019.