SATVIK DIXIT

MS Student in Electrical and Computer Engineering, Carnegie Mellon University
Email: satvikdixit@cmu.edu | Website: https://satvik-dixit.github.io/ | LinkedIn | Google Scholar

EDUCATION

Carnegie Mellon University — Pittsburgh, PA

Master of Science in Electrical and Computer Engineering — Aug 2023 - Dec 2024

- Research areas: Audio and Speech Processing | GPA: 4.0/4.0
- Advisors: Dr Bhiksha Raj, Dr. Chris Donahue

Indian Institute of Technology (IIT) Delhi — New Delhi, India

Bachelor of Technology in Electrical Engineering — Aug 2019 - Aug 2023

- Research areas: ML, Signal Processing | GPA: 8.6/10.0

EXPERIENCE

Carnegie Mellon University

Research Assistant | Professor Bhiksha Raj — May 2024 - Sept 2024

- Developed a novel MFCon (Multi-scale Feature Contrastive) loss, for speaker verification systems, achieving a 9.05% improvement in Equal Error Rate (EER) over SOTA on the VoxCeleb-1O benchmark
- Showed that explicitly enhancing the speaker separability of the intermediate feature maps by using contrastive losses, improves the discriminative ability of the final speaker embedding
- Conducted comprehensive ablation studies to identify optimal configurations and hyperparameters
- Submitted results to ICASSP 2025 [5][PDF][Code]

Massachusetts Institute of Technology

Research Assistant | Professor Satrajit Ghosh — May 2022 - Aug 2023

- Developed a novel framework to explain deep learning embeddings for speech-emotion recognition
- Implemented the probing-based method and evaluated it by explaining WavLM embeddings using EgeMAPS acoustic features for the RAVDESS and SAVEE datasets
- Created a novel metric, Information Increase, to quantify the relevance of specific acoustic features in the embedding and identified the most important feature categories
- Submitted results to ICASSP 2025 [4][PDF][Code]

PROJECTS

Leveraging Audio To Evaluate Audio Captioning Systems

Advisor: Professor Bhiksha Raj, CMU — Sep 2024 - Oct 2024

- Developed a novel metric MACE (Multimodal Audio Caption Evaluation) - the first metric that incorporates both audio and reference captions for comprehensive audio caption evaluation
- Achieved a new SOTA with a 3.28% and 4.36% relative human-preference-match accuracy improvement over the widely-used FENSE metric on Clotho-Eval and AudioCaps-Eval benchmarks
- Submitted to ICASSP Speech and Audio Language Models Workshop 2025 [2][PDF][Code]

Evaluating Visual Language Models on Audio Spectrogram Classification

Advisor: Professor Chris Donahue, CMU — July 2024 - Sep 2024

- Developed a novel task VSC (Visual Spectrogram Classification) to evaluate the ability of Vision Language Models to classify audio using spectrogram images
- Benchmarked zero-shot and few-shot performance of state-of-the-art VLMs (GPT-4o, Claude, and Gemini) on the VSC task and performed ablation studies to optimize spectrogram hyperparameters

- Conducted human studies to show VLMs display human-expert-level performance on the VSC task
- Accepted at Neurips Audio Imagination Workshop 2025 [3][PDF]

Text to Audio Morph Generation

Advisor: Professor Chris Donahue, CMU July 2024 - Present

- Worked on combining two or more categories of sounds to produce novel hybrid sounds
- Implemented methods to extend generative text-to-audio models to be capable of combining text and audio prompts with user defined weights and temporal envelopes

Auditing Audio Text Datasets

Advisor: Professor Chris Donahue, CMU Oct 2024 - Present

- Worked on an audit of 15+ widely-used audio-text datasets
- Evaluated the dataset caption quality by judging factors such as accuracy, coverage, vagueness, etc.

Automatic Speech Recognition For Low Resource Languages

Advisor: Professor Shinji Watanabe, CMU Jan 2024 - Mar 2024

- Added the ASR recipe for a Luganda (African dialect) dataset to the lab's ESPNet toolkit (7k+ stars)
- Achieved a 28.3% improvement in the WER compared to baseline by using CTC-attention-based architecture with spec augment and speed perturbation [PR]

Room Acoustics Simulation

Advisor: Dr. Robin Scheibler, EPFL June 2021 - Aug 2021

- Worked on developing Pyroomacoustics: an open-source package for room acoustics simulation
- Improved RIR simulation accuracy by adding a 'directivity' functionality to mics and sources [demo]

SELECTED PUBLICATIONS

[1] Satvik Dixit,

[2] Satvik Dixit, Soham Deshmukh, Bhiksha Raj. "MACE: Leveraging Audio for Evaluating Audio Captioning Systems." (under review at **ICASSP SALMA** Workshop 2025)

[3] Satvik Dixit, Laurie Heller, Chris Donahue. "Vision Language Models Are Few-Shot Audio Spectrogram Classifiers." **NeurIPS Audio Imagination** Workshop 2024

[4] Satvik Dixit, Massa Baali, Rita Singh, and Bhiksha Raj. "Improving Speaker Representations Using Contrastive Losses on Multi-scale Features." (under review at **ICASSP** 2025)

[5] Satvik Dixit, Daniel Low, Gasser, Fabio Catania, Satrajit Ghosh. "Explaining Deep Learning Embeddings for Speech Emotion Recognition by Predicting Interpretable Acoustic Features." (under review at **ICASSP** 2025)

EXTRACURRICULAR ACTIVITIES

Teaching Assistant: Signals and Systems (18290) for Fall 2024

Teaching Assistant: Signals and Systems (18290) for Spring 2024

Reviewer: ICASSP Speech and Audio Language Models (SALMA) Workshop 2025

SKILLS

Programming Languages: Python, Java, LaTeX, Linux, MATLAB

Frameworks and Tools: PyTorch, Hugging Face, GCP, AWS, Git, CUDA, Speechbrain, ESPNet

CMU Coursework: Speech Recognition and Understanding, Deep Generative Modeling, Advanced Natural Language Processing, Machine Learning (ML), Deep Learning, ML for Signal Processing