SATVIK DIXIT

Email: satvikdixit@cmu.edu | Website: https://satvik-dixit.github.io/ | Google Scholar

EDUCATION

Carnegie Mellon University Pittsburgh, PA
Master of Science in Electrical and Computer Engineering Aug 2023 - Dec 2024
- **Research:** Audio Language Models, Generative Audio | **GPA:** 3.95/4.0
- **Advisors:** Prof. Chris Donahue, Prof. Bhiksha Raj

Indian Institute of Technology (IIT) Delhi New Delhi, India
Bachelor of Technology in Electrical Engineering Aug 2019 - Aug 2023
- **Research:** ML, Signal Processing | **GPA:** 8.7/10.0

RESEARCH EXPERIENCE

Research Assistant | Advisor: Professor Bhiksha Raj, CMU May 2024 - Present

1. Mellow: A Small Audio Language Model For Reasoning [1]
- Developed Mellow, a compact ALM competitive with large-scale models on audio reasoning tasks, trained with 60× less data and 50× fewer parameters
- Conducted extensive ablation studies to identify optimal architectural choices, synthetic data generation methods, and training strategies for creating efficient small ALMs

2. AURA: A Metric for Holistic Audio Question Answering Evaluation
- Built AQEval, the first human-annotated benchmark for Audio QA metrics
- Designed AURA, leveraging LLMs and CLAP to achieve SoTA correlation with human judgments

3. Leveraging Audio To Evaluate Audio Captioning Systems [2]
- Created MACE, the first metric leveraging both audio and reference text for audio caption evaluation
- SoTA by +3.2% (Clotho-Eval) and +4.4% (AudioCaps-Eval) in human preference alignment

4. Improving Speaker Representations Using Contrastive Losses on Multi-scale Features [5]
- Designed a loss function for speaker verification, improving EER by 9.05% on VoxCeleb-1O

Research Assistant | Advisor: Professor Chris Donahue, CMU Aug 2024 - Present

1. Benchmarking Video2Audio models
- Developing a large-scale benchmark for evaluating video-conditioned audio generation models
- Creating an automatic pipeline for converting internet scraped videos into foley-style clips

2. Project: Evaluating Visual Language Models on Audio Spectrogram Classification [4]
- Proposed Visual Spectrogram Classification (VSC) task to evaluate the ability of Large VLMs (such as GPT-4o) to classify audio using spectrogram images alone
- Showed VLMs achieve near human-expert performance in zero/few-shot settings

3. Project: Controllable Audio Morphing [3]
- Developed a framework for combining audio envelopes in a perceptually relevant manner

Summer Research Intern | Advisor: Dr. Satrajit Ghosh, MIT May 2022 - Aug 2023

Explaining DL Embeddings for SER by Predicting Interpretable Acoustic Features [6]
- Worked on interpretability of speech embeddings for speech emotion recognition

Summer Research Intern | Advisor: Dr. Robin Scheibler, EPFL June 2021 - Aug 2021

Acoustics simulation

- Added mic/source directivity support to Pyroomacoustics, a toolkit for indoor acoustics simulation

PUBLICATIONS & PREPRINTS

[1]"Mellow: a small audio language model for reasoning."
Soham Deshmukh, **Satvik Dixit**, Rita Singh, Bhiksha Raj | (under review at **NeurIPS 2025**) [Paper][Code]
[2]"MACE: Leveraging Audio for Evaluating Audio Captioning Systems"
Satvik Dixit, Soham Deshmukh, Bhiksha Raj | **ICASSP SALMA 2025** [Paper][Code]
[3] "Learning Perceptually Relevant Audio Envelope Morphing"
Satvik Dixit, Sungjoon Park, Chris Donahue, Laurie Heller. | **WASPAA 2025** [Paper]
[4]"Vision Language Models Are Few-Shot Audio Spectrogram Classifiers"
Satvik Dixit, Laurie Heller, Chris Donahue | **NeurIPS Audio Imagination 2024** [Paper]
[5]"Improving Speaker Representations Using Contrastive Losses on Multi-scale Features"
Satvik Dixit, Massa Baali, Rita Singh, and Bhiksha Raj | (preprint) [Paper]
[6]"Explaining DL Embeddings for Speech Emotion Recognition by Predicting Interpretable Acoustic Features"
Satvik Dixit, Daniel Low, Gasser, Fabio, Satrajit Ghosh | (preprint) [Paper]

SERVICE

Teaching Assistant: Signals and Systems (18290) for Fall 2024 & Spring 2024 at CMU
Reviewer: ICASSP SALMA 2025, ICML ML4Audio 2025, IEEE Signal Processing Letters 2025

SKILLS

Programming Languages: Python, Java, Bash, MATLAB, LaTeX
Frameworks and Tools: PyTorch, Hugging Face, GCP, AWS, CUDA, SpeechBrain
CMU Coursework: Speech Recognition and Understanding, Deep Generative Modeling, Advanced Natural Language Processing, Machine Learning, Deep Learning, ML for Signal Processing