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:** 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:** ML, Signal Processing | **GPA:** 8.6/10.0

PUBLICATIONS & PREPRINTS

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

RESEARCH EXPERIENCE

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

Project: Mellow: a small Audio Language Model (ALM) for reasoning [1]

- Developed Mellow, a small ALM competitive with SoTA models on reasoning tasks even with 50× less parameters and being trained on 60x less data
- Created ReasonAQA, a novel 1M instance synthetic dataset to enhance audio reasoning in ALMs
- Conducted extensive ablation studies to identify optimal architectural choices, synthetic data generation methods, and training strategies for creating efficient small ALMs

Project: AURA: An LLM-Based Metric for Comprehensive Audio Question Answering (AQA) Evaluation [2]

- Created AQEval, the first human-annotated benchmark for AQA metrics, revealing weak correlation of existing metrics with human judgment
- Developed AURA, an LLM-based metric incorporating audio through CLAP embeddings, achieving SoTA correlation with human judgment

Project: Leveraging Audio To Evaluate Audio Captioning Systems [4]

- Built MACE, the first metric leveraging both audio and reference text for audio caption evaluation
- Achieved SoTA performance, with 3.2% and 4.4% improvements in human preference match accuracy over the widely-used FENSE metric on Clotho-Eval and AudioCaps-Eval benchmarks

Project: Improving Speaker Representations Using Contrastive Losses on Multi-scale Features [6]

- Designed the MFCon loss for speaker verification, improving EER by 9.05% on VoxCeleb-1O
- Showed contrastive learning on intermediate features improves the discriminative ability of the final speaker embedding

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

Project: Evaluating Visual Language Models on Audio Spectrogram Classification [5]

- Developed a novel task VSC (Visual Spectrogram Classification) to evaluate the ability of multimodal models (such as GPT-4o, Claude, and Gemini) to classify audio using spectrogram images alone
- Benchmarked zero-shot and few-shot performance of SoTA models and showed VLMs achieve human-expert level performance on the VSC task

Project: Controllable Audio Morphing [3]

- Developed a framework for combining audio envelopes in a perceptually relevant manner
- Extended text-to-audio models to support morphing sounds based on text prompts and audio envelopes for enhanced user control over the generated audio

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

Project: Explaining DL Embeddings for SER by Predicting Interpretable Acoustic Features [7]

- Worked on interpretability of speech embeddings for speech emotion recognition

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

Project: Acoustics simulation

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

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