# SATVIK DIXIT

satvikdixit7@gmail.com | Webpage | Google Scholar

## EDUCATION

### Carnegie Mellon University — Pittsburgh, PA

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

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

## SELECTED PUBLICATIONS & PREPRINTS

**[1]** "Mellow: a small audio language model for reasoning."
Soham Deshmukh, **Satvik Dixit**, Rita Singh, Bhiksha Raj | **NeurIPS 2025**

**[2]** "MMAU-Pro: A Challenging and Comprehensive Benchmark for Holistic Evaluation of Audio General Intelligence." Sonal Kumar,Simon Sedlacek ... **Satvik Dixit**, et al. | **AAAI 2026**

**[3]** "FoleyBench: A Benchmark For Video-to-Audio Models."
**Satvik Dixit**, Koichi Saito, Zhi Zhong, Yuki Mitsufuji, Chris Donahue | (**preprint under review at ICASSP 2026**)

**[4]** "AURA Score: A Metric For Holistic Audio Question Answering Evaluation."
**Satvik Dixit**, Soham Deshmukh, Bhiksha Raj | (**preprint under review at ICASSP 2026**)

**[5]** "MACE: Leveraging Audio for Evaluating Audio Captioning Systems."
**Satvik Dixit**, Soham Deshmukh, Bhiksha Raj | **ICASSP SALMA 2025**

**[6]** "Learning Perceptually Relevant Temporal Envelope Morphing."
**Satvik Dixit**, Sungjoon Park, Chris Donahue, Laurie Heller | **WASPAA 2025**

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

## EXPERIENCE

### Cekura (YC 24)

**Founding Engineer** — Sept 2025 - Present

- Built an evaluation pipeline for voice agents, simulating naturalistic user-agent conversations across diverse scenarios, user personas, languages, accents, and noise conditions, supporting thousands of automated test conversations per day
- Created a set of granular evaluation metrics for conversational failure-mode analysis including context retention, hallucination detection, instruction following, backchanneling, and interruption handling providing insights into agent reliability

### G-Clef Lab CMU

**Research Assistant | Advisor: Professor Chris Donahue** — Aug 2024 - Aug 2025

- Created **FoleyBench**, the first large-scale benchmark designed for evaluating Foley-style (non-speech/music, causally visual–audio aligned) video-to-audio generation [page][paper]
- Designed a framework for morphing the temporal envelopes of two audio signals [paper]
- Introduced **Visual Spectrogram Classification**, a task evaluating whether VLMs (such as GPT-4o, Gemini) can interpret audio purely through spectrogram images; ran zero-/few-shot evaluations, ablations, and human studies showing VLMs reach expert-level classification accuracy. [paper]

### MLSP Lab CMU

**Research Assistant | Advisor: Professor Bhiksha Raj, CMU** May 2024 - Jan 2025

- Developed **Mellow**, the first small audio language model (167M params) competitive with large-scale models on audio reasoning tasks. Conducted extensive ablation studies to identify optimal architectural choices, synthetic data generation methods, and training strategies for creating efficient small ALMs. Created **ReasonAQA**, a large scale synthetic audio reasoning focused dataset [paper]
- Developed **AURA Score**, a metric for open-ended audio question answering, combining LLM-based reasoning with an audio-entailment check. Demonstrated SotA correlation with human judgments on a new 10k human response benchmark **AQEval**. [paper]
- Created **MACE**, the first metric leveraging audio for audio caption evaluation. SoTA by +3.2% (Clotho-Eval) and +4.4% (AudioCaps-Eval) in human preference alignment [paper]
- Designed a multi-scale contrastive loss for speaker verification, achieved a 9.05% improvement in Equal Error Rate (EER) on the VoxCeleb-1O benchmark [paper]

### Senseable Intelligence Lab MIT

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

- Developed a probing-based framework to explain pre-trained speech representations (such as WavLM) through acoustic features for the speech emotion recognition task [paper]

### LCAV Lab EPFL

**Research intern | Advisor: Dr. Robin Scheibler,** 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]

## SERVICE

**Teaching Assistant**: Signals and Systems (18290) for Fall 2024 & Spring 2024 at CMU
**Reviewer**: ICASSP SALMA 2025, ICML ML4Audio 2025, IEEE SPL 2025, ISMIR LLM4Music 2025
**Mentor**: Mentored 3 graduate students as a CMU Peer Mentor during Spring 2024 and Fall 2024

## SKILLS AND COURSEWORK

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