Shikhar Bharadwaj

PhD Student, **Carnegie Mellon University**

shikhar-s.github.io shikhar.ssu@gmail.com Google Scholar github.com/Shikhar-S

Education

Carnegie Mellon University — Pittsburgh, PA, USA
Advisors: David Mortensen and Shinji Watanabe — *August 2024 - Present*

Indian Institute of Science — Bengaluru, India
M.Tech (Research) in Intelligent Systems — *August 2019 - June 2022*

Birla Institute of Technology and Science — Hyderabad, India
B.E. (Honors) in Computer Science Engineering — *August 2014 - June 2018*

Research Experience

Meta Reality Labs — Redmond, USA
Research Scientist Intern — *May 2026 - August 2026*

Mitsubishi Electric Research Laboratories — Boston, USA
PhD Research Intern — *May 2025 - August 2025*

Google DeepMind — Bengaluru, India
Pre-Doctoral Researcher — *May 2022 - July 2024*

Microsoft Research — Bengaluru, India
Research Intern — *December 2017 - May 2018*

Selected Manuscripts

* denotes equal contribution

[1] **Drowning in Noise: On Signal Saliency in Self-Supervised Learning for Ocean Acoustics**
Shikhar Bharadwaj, Samuele Cornell, Shinji Watanabe, Malcolm Slaney
DCASE 2026 (Best Student Paper Award)

[2] **Phone Segmentation and Recognition through Phonological Activation Mapping**
Shikhar Bharadwaj*, Kwanghee Choi*, Stephen McIntosh*, Chin-Jou Li, Eunjung Yeo, Daisuke Saito, Nobuaki Minematsu, Shinji Watanabe, Jian Zhu, David Harwath, David Mortensen
SLT 2026

[3] **An Empirical Recipe for Universal Phone Recognition**
Shikhar Bharadwaj, Chin-Jou Li, Kwanghee Choi, Eunjung Yeo, William Chen, Shinji Watanabe, David Mortensen
Interspeech 2026

[4] **The CMU-AIST Submission for the ICME 2025 Audio Encoder Challenge**
Shikhar Bharadwaj*, Samuele Cornell*, Kwanghee Choi, Hye-jin Shim, Soham Deshmukh, Satoru Fukayama, Shinji Watanabe
Technical Report 2026

[5] **PRiSM: Benchmarking Phone Realization in Speech Models**
Shikhar Bharadwaj*, Chin-Jou Li*, Yoonjae Kim*, Kwanghee Choi, Eunjung Yeo, Ryan Shim, Hanyu Zhou, Brendon Boldt, Karen Jacome, Kalvin Chang, Darsh Agrawal, Keer Xu, Huck Yang, Jian Zhu, Shinji Watanabe, David Mortensen
ACL 2026

[6] **POWSM: A Phonetic Open Whisper-Style Speech Foundation Model**
Chin-Jou Li*, Kalvin Chang*, Shikhar Bharadwaj, Eunjung Yeo, Kwanghee Choi, Jian Zhu, David Mortensen, Shinji Watanabe
ACL 2026 (Oral)

[7] **OpenBEATs: A Fully Open-Source General-Purpose Audio Encoder**
Shikhar Bharadwaj, Samuele Cornell, Kwanghee Choi, Satoru Fukayama, Hye-jin Shim, Soham Deshmukh,

Shinji Watanabe
WASPAA 2025 (Oral)

[8] **Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities**
Gemini Team, Google (including Shikhar Bharadwaj)
Technical Report 2025

[9] **OpusLM: A Family of Open Unified Speech Language Models**
Jinchuan Tian, William Chen, Yifan Peng, Jiatong Shi, Siddhant Arora, Shikhar Bharadwaj, Takashi Maekaku, Yusuke Shinohara, Keita Goto, Xiang Yue, Huck Yang, Shinji Watanabe
Interspeech 2025

[10] **IndicGenBench: A Multilingual Benchmark to Evaluate Generation Capabilities of LLMs on Indic Languages**
Harman Singh, Nitish Gupta, Shikhar Bharadwaj, Dinesh Tewari, Partha Talukdar
ACL 2024

[11] **STAB: Speech Tokenizer Assessment Benchmark**
Shikhar Vashishth*, Harman Singh*, Shikhar Bharadwaj*, Sriram Ganapathy, Chulayuth Asawaroengchai, Kartik Audhkhasi, Andrew Rosenberg, Ankur Bapna, Bhuvana Ramabhadran
Preprint 2024

[12] **Multimodal Modeling for Spoken Language Identification**
Shikhar Bharadwaj*, Min Ma*, Shikhar Vashishth*, Ankur Bapna, Sriram Ganapathy, Vera Axelrod, Siddharth Dalmia, Wei Han, Yu Zhang, Daan van Esch, Sandy Ritchie, Partha Talukdar, Jason Riesa
ICASSP 2024

Selected Research Directions

- **Multilingual Speech Representations**
  - Trained PhoneticXEUS [3] , a state-of-the-art universal phone recognition model.
  - Created PRiSM [5] to benchmark the phonetic capabilities of speech models, focusing on phone recognition in unseen languages and downstream low-resource application areas such as accented, multilingual, and pathological speech.
  - Developed POWSM [6], a phonetic open Whisper-style speech foundation model that unifies multiple phone-related tasks including ASR, phone recognition, grapheme-to-phoneme and phoneme-to-grapheme.
  - Created MuSeLi [12] to jointly model speech and text for language identification at the scale of YouTube. This is the current SOTA for language identification and received several awards at Google.
- **Audio Representation Learning**
  - Trained OpenBEATs [7] , a masked audio token prediction based audio encoder.
  - By scaling pre-training to 70K hours of open-source speech and audio, OpenBEATs outperforms billion-parameter MAE models using only 25% of their parameter count.
  - Our submission based on OpenBEATs was the top performing self-supervised model in the ICME 2025 Audio Encoder Challenge [4], and achieved third place overall.
  - Studied signal saliency in SSL pre-training [1] , showing that 30 hours of salient ocean audio outperform 1,240 hours of unfiltered data.

Awards and Recognition

- **Best Student Paper Award**: DCASE 2026 Workshop [1]
- **Google Internal Awards**: Five awards for contributions towards Google4India event, Google-USM and multimodal language identification work [12]
- **Media Coverage**: Audio collection effort in collaboration with the *Indian Institute of Science* has been covered by The White House and The Economic Times
- **Academic Excellence**: Institute Merit Scholarship (Top 2% for 7 semesters, Top 1% for 1 semester), NTSE Scholarship, INSPIRE Scholarship