VISHAL SUNDER

(+1) 6149725371 ◇ sunder.9@osu.edu ◇ https://vishalsunder.github.io

RESEARCH INTERESTS

Automatic Speech Recognition, Speech Understanding, Speech Synthesis

EDUCATION

PhD candidate in CSE *December 2024 (expected)*
The Ohio State University *Overall CPI: 3.9/4.00*
Advisor: Dr. Eric Fosler-Lussier

Bachelor of Technology in Electrical Engineering *May 2016*
Indian Institute of Technology (BHU), Varanasi *Overall CPI: 8.35/10*

PROFESSIONAL EXPERIENCE

Graduate Research Associate *Present*
The Ohio State University
Speech and Language Technologies Lab

Research Intern *May 2024 - August 2024*
IBM Research, Yorktown Heights, USA
Speech Technologies group

Research Intern *May 2022 - August 2022*
IBM Research, Yorktown Heights, USA
Speech Technologies group

Research Intern *May 2021 - August 2021*
IBM Research, Yorktown Heights, USA
Speech Technologies group

Research Engineer *July 2016 - July 2019*
TCS Research, New Delhi, India
Deep Learning and Artificial Intelligence group

PUBLICATIONS

V. Sunder, G. Saon, S. Thomas, B. Kingsbury, S. Shechtman, E. Fosler-Lussier, L. Lastras. A Non-autoregressive Multimodal Approach for Joint STT and TTS. *Under preparation for ICASSP 2025*

V. Sunder, E. Fosler-Lussier. Improving Transducer-based Spoken Language Understanding with Self-conditioned CTC and Knowledge Transfer. *Under review at IEEE Spoken Language Technology Workshop (SLT), 2024*

V. Sunder, B. Karrolla, E. Fosler-Lussier. End-to-End real time tracking of children's reading with pointer network. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024* [Paper]

V. Sunder, E. Fosler-Lussier, S. Thomas, HKJ. Kuo, B. Kingsbury. ConvKT: Conversation-Level Knowledge Transfer for Context Aware End-to-End Spoken Language Understanding. *INTERSPEECH-2023.*[Paper]

V. Sunder, S. Thomas, HKJ. Kuo, B. Kingsbury, E. Fosler-Lussier. Fine-grained Textual Knowledge Transfer to Improve RNN Transducers for Speech Recognition and Understanding. *ICASSP-2023.*[Paper]

L. Venkatasubramaniam*, **V. Sunder***, E. Fosler-Lussier. End-to-End word-level disfluency detection and classification in children's reading assessment. *ICASSP-2023.*[Paper] [Code]

K. R Maicher, A. Stiff, M. Scholl, M. White, E. Fosler-Lussier, W. Schuler, P. Serai, **V. Sunder**, H. Forrestal, L. Mendella, M. Adib, C. Bratton, K. Lee, D. R. Danforth. Artificial intelligence in virtual standardized patients: Combining natural language understanding and rule based dialogue management to improve conversational fidelity. *Medical Teacher, Volume 45.* [Paper]

V. Sunder, E. Fosler-Lussier, S. Thomas, HKJ. Kuo, B. Kingsbury. Tokenwise Contrastive Pretraining for Finer Speech-to-BERT Alignment in End-to-End Speech-to-Intent Systems. *INTERSPEECH-2022.* [Paper] [Code]

V. Sunder, S. Thomas, HKJ. Kuo, J. Ganhotra, B. Kingsbury, E. Fosler-Lussier. Towards End-to-End Integration of Dialog History for Improved Spoken Language Understanding. *ICASSP-2022.* [Paper]

P. Serai, **V. Sunder**, E. Fosler-Lussier. Hallucination of speech recognition errors with sequence to sequence learning. *IEEE/ACM Transactions on Audio, Speech and Language Processing.* [Paper]

V. Sunder, P. Serai, E. Fosler-Lussier. Building an ASR Error Robust Spoken Virtual Patient System in a Highly Class-Imbalanced Scenario Without Speech Data. [Paper]

V. Sunder, E. Fosler-Lussier. Handling Class Imbalance in Low-Resource Dialogue Systems by Combining Few-Shot Classification and Interpolation. *ICASSP 2021.* [Paper] [Code]

V. Sunder, A. Srinivasan, L. Vig, G. Shroff, R. Rahul: One-shot information extraction from document images using neuro-deductive program synthesis. *NeSy workshop, IJCAI 2019.* [Paper]

G. Gupta,**V. Sunder**, R. Prasad, G. Shroff. CRESA: A Deep Learning Approach to Competing Risk Recurrent Event Survival Analysis. *PAKDD-2019.* [Paper]

V. Sunder, L. Vig, A. Chatterjee, G. Shroff. Prosocial or Selfish? Agents with different behaviors for Contract Negotiation using Reinforcement Learning. *ACAN workshop, IJCAI 2018.* [Paper]

V. Sunder, M. Yadav, L. Vig, G. Shroff. Information Bottleneck Inspired Method for Chat Text Segmentation. *IJCNLP 2017.* [Paper]

PATENTS

V. Sunder, L. Vig, A. Chatterjee, G. Shroff. Method and system for performing negotiation task using reinforcement learning agents. *US Patent 11,521,281.* [Link]

S. Thomas, **V. Sunder**, H.K Kuo, B. Kingsbury, E. Fosler-Lussier, G. Saon. Textual knowledge transfer for speech recognition and understanding. *Applied US Patent 18/310,598*

S. Thomas, **V. Sunder**, H.K Kuo, J. Ganhotra, B. Kingsbury, E. Fosler-Lussier. End-to-End integration of dialog history for spoken language understanding. *Applied US Patent 17/665,441*

SELECTED RESEARCH PROJECTS

Combining speech and language foundational models for semantic understanding of speech @ OSU [August 2024 - present]

- Developing tools for combining various speech foundation models and LLMs so they can work in synergy for complex speech understanding tasks.

- Exploring various interfacing models like Q-former, CTC-compression etc. to effectively align speech and language models.

Non-autoregressive multimodal models for joint STT and TTS @ IBM [May 2024 - Aug 2024]

- Built multimodal conformer models, trained jointly with speech-to-text and text-to-speech tasks.
- Model is completely non-autoregressive allowing for fast inference for both STT and TTS.
- Defined various modality-to-modality (self-)supervised tasks leading to transfer learning and an iterative refinement decoding algorithm.
- Work being prepared for ICASSP-2025.

Deep Learning models for children's reading assessment @ OSU [May 2022 - Aug 2023]

- Built an end-to-end disfluency detection and classification system for children read speech using a novel hierarchical training framework.
- Built an end-to-end real-time tracking system for children read speech using pointer networks.
- Work published at ICASSP-2023, ICASSP-2024.

Robust end-to-end speech understanding @ OSU & IBM Research [May 2021 - present]

- Built a fully end-to-end speech understanding system to integrate dialog history in speech form.
- Designed a learning paradigm to align acoustic and LLM embeddings at the token level.
- Ongoing work on integrating speech encoders and LLM layers to facilitate end-to-end learning.
- Work published at ICASSP-2022, INTERSPEECH-2022, ICASSP-2023, INTERSPEECH-2023.

The Virtual Patient project @ OSU [May 2020 - April 2021]

- Developed a novel pairwise training framework for handling long-tailed class imbalance issue in a domain specific question classification dataset.
- Built an ASR error robust NLU system by utilizing an ASR error simulation framework.
- Work published at ICASSP-2021 and the TASLP journal.

TECHNICAL SKILLS

Languages: Python, **DL frameworks:** PyTorch, TorchAudio, SpeechBrain

Experience in building RNN-T, AED and CTC based ASR models in PyTorch [Link]

ACADEMIC SERVICE

Reviewer - ICLR 2023, NeurIPS 2024, TASLP 2024