Natarajan Balaji Shankar

310-254-0577 balaji1312@ucla.edu balaji1312 balaji1312.github.io balaji1312

Education

University of California, Los Angeles (UCLA) **Expected, June, 2026**
Ph.D, Electrical and Computer Engineering (GPA: 4.00/4.00) *Los Angeles, CA*

University of California, Los Angeles (UCLA) **2023**
M.S., Electrical and Computer Engineering (GPA: 3.97/4.00) *Los Angeles, CA*

National Institute of Technology, Tiruchirappalli (NIT Trichy) **2020**
B.Tech, Electronics and Communication Engineering (Minor: Computer Science) (GPA: 8.57/10.00) *Tiruchirappalli, India*

Technical Skills

Languages: Python, C, C++, MATLAB
Frameworks: PyTorch, Tensorflow, Kaldi, LangChain, Huggingface, Scikit-learn, NLTK, Flask
Developer Tools: Linux, PostgreSQL, Docker, Git

Research Projects

Low Resource ASR for Child Speech

- Introduced novel Speech-Only Adaptation (SOA) technique for unsupervised domain adaptation of speech foundation models to low resource domains utilizing only audio data, resulting in a 29% relative Word Error Rate reduction
- Proposed Perturbation Invariant Finetuning (PIF) loss to stabilize finetuning on augmented child speech, resulting in state-of-the-art results on the MyST dataset
- Designed a comprehensive benchmark to evaluate child ASR performance across various speech foundation models (Whisper, Wav2vec2.0, HuBERT, WavLM), while investigating finetuning strategies (data augmentation, parameter-efficient methods) and their impact based on model size

Automated Scoring and Analysis of Children's Oral Language Assessments

- Developed novel framework for automatically scoring children's oral narrative language abilities using linguistic features and BERT-based representations, achieving 98% classification accuracy of student scores on a portion of the Test of Narrative Language
- Evaluated the performance of open-source ASR systems (Whisper, HuBERT, Wav2Vec2) in transcribing African American English-speaking children's speech and used BERT to automatically grade oral responses with 95% accuracy
- Incorporated a pipeline for automated prediction of item difficulty and response time in educational assessments using NER, SRL, and linguistic features to prompt a Large Language Model, yielding results that surpassed the baseline on the BEA 2024 Shared Task and showcasing adaptability on Test of Narrative Language audio recordings

Encoder Only Non-autoregressive ASR

- Devised UniEnc-CASSNAT, a novel encoder-based non-autoregressive automatic speech recognition (NASR) model combining the strengths of CTC and Attention decoding
- Leveraged a single encoder for mask prediction and text decoding, enabling efficient integration of speech foundation models
- Produced state-of-the-art results on Librispeech, MyST, and Aishell1 datasets with fewer model parameters than existing NASR models

CORAAL QA: A database and framework for question answering from long audio files

- Created a novel dataset (CORAAL QA) and framework for audio question-answering from long audio recordings of spontaneous speech
- Utilized PLDA scoring of BERT-based semantic embeddings to rank ASR transcript segment relevance for answering questions
- Employed large language models (ChatGPT, Llama 2) to generate additional training examples, enhancing framework performance through data augmentation

Predicting Dialect Density in African American English

- Formulated novel framework for predicting dialect density in African American English (AAE) speakers, demonstrating high classification accuracy across age and regional variations
- Integrated usage of HuBERT representations, prosody, grammar, speaker embedding based feature sets to train a classifier for accurate dialect density labeling
- Enhanced specificity in downstream tasks by predicting the level of dialect density rather than relying on a binary presence/absence approach, with a 40% accuracy improvement over baselines in adults and over 70% accuracy in children

Experience

Speech Processing and Auditory Perception Laboratory **01/2022 - 09/2024**

Graduate Student Researcher *Los Angeles, CA*

- Instituted comprehensive benchmarks and finetuning strategies for supervised and unsupervised child ASR performance evaluation on speech foundation models
- Devised frameworks for automatically scoring children's oral language abilities, achieving high accuracy in predicting scores from various tasks, including narrative language, picture descriptions, and responses in African American English
- Developed UniEnc-CASSNAT, a state-of-the-art encoder-based non-autoregressive ASR model
- Introduced CORAAL QA, a novel dataset and framework for question-answering from long audio files, utilizing PLDA scoring and large language models for enhanced performance

KLA Corporation **06/2022 - 09/2022**

Algorithms Intern *Milpitas, CA*

- Migrated backend for wafer inspection tool from Windows to Linux to facilitate multi GPU execution
- Created data transmission framework to enable client side wafer inspection tool GUI to communicate with Linux based remote backend
- Constructed a license generation system using Python and Flask to authenticate access requests for eligible clients for an internal tool

Publications

- **N. B. Shankar**, A. Johnson, C. Chance, H. Veeramani, and A. Alwan. *CORAAL QA: A Dataset and Framework for Open Domain Spontaneous Speech Question Answering from Long Audio Files.* ICASSP 2024 - IEEE International Conference on Acoustics, Speech, and Signal Processing (2024)
- **N. B. Shankar**, R. Fan, and A. Alwan. *SOA: Reducing Domain Mismatch in SSL Pipeline by Speech Only Adaptation for Low Resource ASR.* 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (2024)
- R. Fan, **N. B. Shankar**, and A. Alwan. *Benchmarking Children's ASR with Supervised and Self-supervised Speech Foundation Models.* Proc. Interspeech (2024)
- R. Fan, **N. B. Shankar**, and A. Alwan. *UniEnc-CASSNAT: An Encoder-only Non-autoregressive ASR for Speech SSL Models.* IEEE Signal Processing Letters (2024)
- A. Johnson, **N. B. Shankar**, M. Ostendorf, and A. Alwan. *An Exploratory Study on Dialect Density Estimation for Children and Adult's African American English.* The Journal of the Acoustical Society of America (2024)
- H. Veeramani, S. Thapa, **N. B. Shankar**, and A. Alwan. *Large Language Model-based Pipeline for Item Difficulty and Response Time Estimation for Educational Assessments.* Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA) (2024)
- A. Johnson, H. Veeramani, **N. B. Shankar**, and A. Alwan. *An Equitable Framework for Automatically Assessing Children's Oral Narrative Language Abilities.* Proc. Interspeech (2023)
- A. Johnson, C. Chance, K. Stiemke, H. Veeramani, **N. B. Shankar**, and A. Alwan. *An Analysis of Large Language Models for African American English Speaking Children's Oral Language Assessment.* Journal of Black Excellence in Engineering, Science, and Technology (2023)
- H. Veeramani, A. Johnson, **N. B. Shankar**, and A. Alwan. *Towards Automatically Assessing Children's Oral Picture Description Tasks.* Proc. 9th Workshop on Speech and Language Technology in Education (SLaTE) (2023)
- G. Palanisamy, **N. B. Shankar**, P. Palanisamy, and V. P. Gopi. *A hybrid feature preservation technique based on luminosity and edge based contrast enhancement in color fundus images.* Biocybernetics and Biomedical Engineering (2020)