NATARAJAN BALAJI SHANKAR

📞 310-254-0577 ✉ balaji1312@ucla.edu 💼 balaji1312 🌐 balaji1312.github.io ⭕ balaji1312

Education

University of California, Los Angeles (UCLA)   **2023 - Present**
Doctor of Philosophy, Electrical and Computer Engineering   *Los Angeles, CA*

University of California, Los Angeles (UCLA)   **2021 - 2023**
Master of Science, Electrical and Computer Engineering   *Los Angeles, CA*

National Institute of Technology, Tiruchirappalli (NIT Trichy)   **2016 - 2020**
Bachelor of Technology, Electronics and Communication Engineering (Minor: Computer Science)   *Tiruchirappalli, India*

Technical Skills

Languages: Python, C, C++, MATLAB
Frameworks: PyTorch, Tensorflow, Kaldi, LangChain, Huggingface, fairseq, NeMo, Espnet, NLTK
Developer Tools: Linux, PostgreSQL, Docker, Git

Experience

Speech Processing and Auditory Perception Laboratory   **01/2022 - Present**
Graduate Student Researcher   *Los Angeles, CA*

Model Merging for Low-Resource Automatic Speech Recognition
- Developed Selective Attention (SA) Merge, a novel approach that merges task vectors from attention matrices, enhancing Speech Foundation Model (SFM) performance for low-resource tasks such as child Automatic Speech Recognition.
- Achieved a 14% relative Word Error Rate (WER) reduction and a state-of-the-art WER of 8.69 on the MyST database for the Whisper-small model by combining SA Merge with data augmentation techniques.

Child ASR Hypothesis Error Correction
- Conducted comprehensive analyses of transcription errors made by Automatic Speech Recognition (ASR) models on child speech, identifying error patterns linked to acoustic and linguistic variations.
- Investigated correction frameworks leveraging Large Language Models (LLMs) and speech-LLMs, developing approaches for acoustically conditioned child speech error correction.

Improving Child Automatic Speech Recognition
- 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 WER reduction.
- Proposed Perturbation Invariant Finetuning (PIF) loss to stabilize finetuning on augmented child speech, achieving state-of-the-art results on the MyST dataset.

Automated Child Language Assessment
- Developed framework for automatically scoring children's oral narrative language abilities using linguistic features and BERT-based representations, achieving 98% classification accuracy on a portion of the Test of Narrative Language.
- Evaluated open-source ASR systems (Whisper, HuBERT, Wav2Vec2) on African American English-speaking children's speech and used BERT to grade oral responses with 95% accuracy.

Encoder-Only Model for Non-autoregressive Speech Recognition
- Devised UniEnc-CASSNAT, a novel encoder-based non-autoregressive ASR model combining CTC and attention decoding, leveraging a single encoder for mask prediction and text decoding.
- Produced state-of-the-art results on Librispeech, MyST, and Aishell1 datasets with fewer model parameters than existing NASR models.

Question Answering from Long Context Audio
- Created a novel dataset (CORAAL QA) and framework for audio question-answering from long audio recordings of spontaneous speech, utilizing PLDA scoring of semantic embeddings to rank 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 framework for predicting dialect density in African American English speakers integrating HuBERT representations, prosody, grammar, and speaker embedding-based feature sets.

- Enhanced specificity in downstream tasks by predicting the level of dialect density, achieving a 40% accuracy improvement over baselines in adults and over 70% accuracy in children.

KLA Corporation   06/2022 - 09/2022
Algorithms Intern   *Milpitas, CA*

- Modernized a wafer inspection tool by migrating its backend to Linux, enabling multi-GPU execution and boosting performance.
- Engineered a data transmission framework for a wafer inspection tool GUI to seamlessly communicate with a Linux-based backend.
- Developed a secure Python/Flask license generation system to authenticate client access for an internal tool.

Academic and Research Contributions

- Served as a peer reviewer for the IEEE/ACM Transactions on Audio, Speech, and Language Processing, IEEE Signal Processing Letters, International Joint Conference on Neural Networks (IJCNN), and ICASSP 2025 - IEEE International Conference on Acoustics, Speech, and Signal Processing.
- Delivered an oral presentation on research findings related to "*CORAAL QA: A Dataset and Framework for Open Domain Spontaneous Speech Question Answering from Long Audio Files*" at the ICASSP 2024 conference.
- Successfully mentored a diverse team of Masters' (6) and undergraduate (5) students across various research projects, leading to successful thesis and project completions.
- Gained four years of teaching experience as a Graduate Teaching Assistant across a variety of courses

Publications

- **Shankar, N. B.**, Wang, Z., Eren, E., & Alwan, A. (2024). Selective Attention Merging for low resource tasks: A case study of Child ASR. *ICASSP 2025 - IEEE International Conference on Acoustics, Speech, and Signal Processing*
- **Shankar, N. B.**, Johnson, A., Chance, C., Veeramani, H., & Alwan, A. (2024). 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*, 13371–13375.
- **Shankar, N. B.**, Fan, R., & Alwan, A. (2024). 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*, 560–564.
- **Shankar, N. B.**, Afshan, A., Johnson, A., Mahapatra, A., Martin, A., Ni, H., Won Park, H., Quintero Perez, M., Yeung, G., Bailey, A., Breazeal, C., & Alwan, A. (2024). The JIBO Kids Corpus: A speech dataset of child-robot interactions in a classroom environment. *JASA Express Lett.* 1 November 2024; 4 (11): 115201.
- Fan, R., **Shankar, N. B.**, & Alwan, A. (2024). Benchmarking Children's ASR with Supervised and Self-supervised Speech Foundation Models. *Proc. Interspeech*, 5173–5177.
- Fan, R., **Shankar, N. B.**, & Alwan, A. (2024). UniEnc-CASSNAT: An Encoder-only Non-autoregressive ASR for Speech SSL Models. *IEEE Signal Processing Letters*, 31, 711–715.
- Johnson, A., **Shankar, N. B.**, Ostendorf, M., & Alwan, A. (2024). An Exploratory Study on Dialect Density Estimation for Children and Adult's African American English. *The Journal of the Acoustical Society of America*, 155(4), 2836–2848.
- Bailey, A., Johnson, A., **Shankar, N. B.**, Veeramani, H., Washington, J., & Alwan, A. (2025). Addressing Bias in Spoken Language Systems Used in the Development and Implementation of Automated Child Language-Based Assessment. *Journal of Educational Measurement*
- Veeramani, H., Thapa, S., **Shankar, N. B.**, & Alwan, A. (2024). 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)*, 561–566.
- Johnson, A., Veeramani, H., **Shankar, N. B.**, & Alwan, A. (2023). An Equitable Framework for Automatically Assessing Children's Oral Narrative Language Abilities. *Proc. Interspeech*, 4608–4612.
- Johnson, A., Chance, C., Stiemke, K., Veeramani, H., **Shankar, N. B.**, & Alwan, A. (2023). 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*, 1.
- Veeramani, H., Johnson, A., **Shankar, N. B.**, & Alwan, A. (2023). Towards Automatically Assessing Children's Oral Picture Description Tasks. *Proc. 9th Workshop on Speech and Language Technology in Education (SLaTE)*, 119–120.
- Palanisamy, G., **Shankar, N. B.**, Palanisamy, P., & Gopi, V. P. (2020). A hybrid feature preservation technique based on luminosity and edge based contrast enhancement in color fundus images. *Biocybernetics and Biomedical Engineering*, 40(2), 752–763.