


Khandokar Md. Nayem

 [nayem.github.io](https://github.com/nayem)

 +1 (812) 369-6358

 knayem@iu.edu

TECHNICAL EXPERTISE

• Deep Learning • Machine Learning • Large Language Model (LLM) • Natural Language Processing (NLP) • Sequence-to-sequence modeling • Digital Signal Processing (DSP) • Speech Enhancement, Translation & Recognition • Classification & Regression • Knowledge Distillation • Machine Learning Models (Bayesian Network, HMM, GMM, Clustering, Decision Tree, Ensemble Methods) • Deep Neural Network (DNN) • Long-Short Term Memory (LSTM) • Convolutional Neural Network (CNN) • Transformers • Deep Learning Libraries (PyTorch, TensorFlow, Keras, Fairseq, HuggingFace) • NLP Libraries (NLTK, Scikit-Learn) • Human Survey Platform (Qualtrics, Amazon MTurk) • Computer Vision & Graphics Libraries (CImg, OpenCV, OpenGL) • Docker • Git • AWS • Sagemaker • Python • Matlab • C/C++ • Java • Python Libraries (Pandas, Numpy) • HTML, CSS & Javascript • PHP • SQL • R • Shell

EMPLOYMENT

Amazon, Seattle, WA, *Applied Scientist II Intern, Seller Partner Services (SPS)* Summer 2023

- Conducted research on the application of the large language model (LLM) for class labeling on closed taxonomy utilizing product descriptions, while also generating chain-of-reasoning explanations for improved overall comprehension.

Amazon, Cambridge, MA, *Applied Scientist II Intern, Alexa AI* Fall 2022

- Researched the development of a real-time, end-to-end compressed multi-lingual speech translation system. Investigated the use of large language models (LLMs) and applied knowledge distillation approach to transfer their performance to smaller models with 50% and 75% fewer parameters. ([INTERSPEECH 2023](#), [poster](#))

Microsoft Research, Redmond, WA, *Audio & Acoustics Research Intern* Summer 2022

- Focused on analyzing and improving the performance of speech enhancement algorithms to generate high-fidelity (Hi-Fi) speech by removing distortions and extending speech bandwidth. Applied causal LSTM models with various augmentation to recover codec and clipping distortions, and performed deep noise suppression.

Bose Corporation, Boston, MA, *Machine Learning/Neural Signal Processing Intern* Summer 2020

- Researched on enhancing speech in remote microphone applications by removing self-speech in order to provide better quality sound with low latency to hearing aids and voice-assistive wearable devices. Utilized an LSTM-based architecture with speaker-dependent d-vector for speaker identification, to ensure real-time operation.

Indiana University, Bloomington, IN, *Research Assistant, ASPIRE research lab* Fall 2016 - Present

- Developed an attention-based monaural speech enhancement model with the objective of maximizing human perceptual rating of enhanced speech. This was accomplished by incorporating embedding vectors from a human Mean-Opinion Score (MOS) prediction model and jointly training the model utilizing real-world noisy speech data. ([INTERSPEECH-2021](#), [arXiv version](#), extended version [TASLP 2024](#))
- Proposed & implemented a quantized speech prediction model that classifies speech spectra into a corresponding quantized class, and applies a language-style model to generate more realistic speech. Acceptable quantization level was determined by listener study conducted on Amazon MTurk, designed using Qualtrics. ([ICASSP-2021](#), [poster](#), [slides](#), [video](#))
- Designed a recurrent layer, named Intra-Spectral Recurrent (ISR) layer to capture spectral dependencies within the magnitude and phase responses of noisy speech using Markovian recurrent connections. This was successfully integrated into a LSTM-based single-channel speech enhancement model. ([ICASSP-2020](#), [slides](#), [video](#))
- Formulated a new type of recurrent output layer that enforces spectral-level dependencies within each spectral time frame, by modeling the Markovian assumption along the frequency axis in both uni-directional and bi-directional ways. This was tested in a magnitude speech enhancement model. ([MLSP-2019](#), [poster](#))
- Engineered a deep architecture, named Recurrent Stacked Generative Adversarial Network (RSGAN) to generate video clips based on a precondition, such as a sentence description, action classes, or fMRI signals. ([IU-VISION-2017](#), [poster](#))

EDUCATION

Ph.D. in Computer Science, December 2023
[Indiana University](#), Bloomington, IN, USA Advisor: [David J. Crandall](#), [Donald S. Williamson](#)

M.Sc. in Computer Science, December 2019
[Indiana University](#), Bloomington, IN, USA

B.Sc. in Computer Science & Engineering (CSE), July 2014
[Bangladesh University of Engineering & Technology \(BUET\)](#), Dhaka, Bangladesh