

# Khandokar Md. Nayem

Luddy Hall  
700 N Woodlawn Ave, #3061P  
Bloomington, IN 47408

+1 (812) 369-6358

[knayem@iu.edu](mailto:knayem@iu.edu)

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



## RESEARCH INTEREST

---

- Deep Learning • Applied Machine Learning • Natural Language Processing (NLP) • Digital Signal Processing (DSP)
- Speech Enhancement, Translation & Recognition • Digital Health • Wearable Computing

## EDUCATION

---

### Ph.D. in Computer Science

[Indiana University](#), Bloomington, IN, USA

Fall 2023 (Anticipated)

Advisor: [Prof. Donald S. Williamson](#)

### M.Sc. in Computer Science

[Indiana University](#), Bloomington, IN, USA

December 2019

### B.Sc. in Computer Science & Engineering (CSE)

[Bangladesh University of Engineering & Technology \(BUET\)](#), Dhaka, Bangladesh

July 2014

## PUBLICATIONS

---

### In-review (Journal & Conference):

- **Khandokar Md. Nayem**, Donald S. Williamson, “Attention-based Speech Enhancement Using Human Quality Perception Modelling”, in Proceedings of journal IEEE/ACM [Transactions on Audio Speech and Language Processing \(TASLP\)](#) 2023.

**Journal Impact factor: 3.919, SCImago Journal Rank (SJR): 0.92**

- **Khandokar Md. Nayem**, Ran Xue, Ching-Yun Chang, Akshaya Vishnu Kudlu Shanbhogue, “Knowledge Distillation on Joint Task End-to-End Speech Translation”, in Proceedings of the Annual Conference of the [International Speech Communication Association \(ISCA\)](#), INTERSPEECH 2023)

**Conference Impact factor: 5.14, SCImago Journal Rank (SJR): 0.689**

### Peer-reviewed & Published (Conference):

- **Khandokar Md. Nayem**, Donald S. Williamson, “Incorporating Embedding Vectors from a Human Mean-Opinion Score Prediction Model for Monaural Speech Enhancement”, in Proceedings of the Annual Conference of the [International Speech Communication Association \(ISCA\)](#), INTERSPEECH 2021. ([paper](#), [slides](#), [video](#))

**Conference Impact factor: 5.14, SCImago Journal Rank (SJR): 0.689**

**Paper Citation in Google Scholar: 4**

- **Khandokar Md. Nayem**, Donald S. Williamson, “Towards an ASR approach using Acoustic and Language Models for Speech Enhancement”, in Proceedings of IEEE [International Conference on Acoustics, Speech and Signal Processing \(ICASSP\)](#) 2021. ([paper](#), [poster](#), [slides](#), [video](#))

**Conference Impact factor: 4.65, SCImago Journal Rank (SJR): 0.546**

**Paper Citation in Google Scholar: 2**

- **Khandokar Md. Nayem**, Donald S. Williamson, “Monaural speech enhancement using intra-spectral recurrent layers in the magnitude and phase responses”, in Proceedings of IEEE [International Conference on Acoustics, Speech and Signal Processing \(ICASSP\)](#) 2020. ([paper](#), [slides](#), [video](#))

**Conference Impact factor: 4.65, SCImago Journal Rank (SJR): 0.546**

**Paper Citation in Google Scholar: 1**

- **Khandokar Md. Nayem**, Donald S. Williamson, “Incorporating Intra-Spectral Dependencies with a Recurrent Output Layer for Improved Speech Enhancement”, in Proceedings of IEEE [International Workshop on Machine Learning for Signal Processing \(MLSP\)](#) 2019. ([paper](#), [poster](#))

**Conference Impact factor: 1.81, SCImago Journal Rank (SJR): 0.283**

**Paper Citation in Google Scholar: 6**

## Poster:

- Shujon Naha, **Khandokar Md. Nayem**, Md. Lisul Islam, “RSGAN: Recurrent Stacked Generative Adversarial Network for Conditional Video Generation”, presented at IU computer vision project showcase, 2017. ([paper](#), [poster](#))
- **Khandokar Md. Nayem**, Mir Toornaw Islam, Md. Monirul Islam, “Handwritten Writer Independent Bangla Character Recognition”, presented at undergrad thesis at BUET, 2014. ([paper](#), [poster](#))

## Archive:

- **Khandokar Md. Nayem**, Donald S. Williamson, “Attention-based Speech Enhancement Using Human Quality Perception Modelling”, in arxiv, 2023. ([paper](#))

## RESEARCH EXPERIENCE

---

### Indiana University, Bloomington, IN, USA

Summer 2018 - Now

*Research Assistant, [ASPIRE research lab](#)*

- Developed a novel [speech enhancement](#) model that aims to maximize human perceptual satisfaction of speech after cleaning real-world noise from the captured audio. We incorporate perceptually important features learned from a separate perception rating prediction model in achieving this task.
- 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 subjective study conducted on [Amazon MTurk](#), designed using [Qualtrics](#).
- Formulated a new type of recurrent output layer in the context of deep learning architecture that shows significant improvement in [speech enhancement](#). This new design utilizes the internal relations along the frequency axis between speech frequency points.
- Engineered a deep architecture named Recurrent Stacked Generative Adversarial Network ([RSGAN](#)) which [generates video](#) clips given different types of signal like human's brain fMRI signal or sentence description of the video clips.
- Developing a [smart system](#) that tracks the daily physical activity and sleep activity of pregnant women collected by wearable devices and helps to diagnose gestational complications, like gestational diabetics and pre-eclampsia. This project is part of the ongoing [NSF Proactive Health Informatics \(PHI\)](#) project.
- Researched on [speech emotion](#) detection systems that analyze speech to monitor human emotions, a valuable cue for tackling sensitive emotional situations and maintaining a healthy conversation.

### Amazon Services LLC, Cambridge, MA, USA

Fall 2022

*Applied Scientist Intern, [Alexa AI](#)*

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

### Microsoft Corporation, Redmond, WA, USA

Summer 2022

*Research Intern, [Audio and Acoustics Research Group](#)*

- 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 real-time [deep learning](#) models with various data augmentation techniques to recover codec and clipping distortions, and performed deep noise suppression.

### BOSE Corporation, Boston, MA, USA

Summer 2020

*Machine Learning/Neural Signal Processing Intern*

- 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 a [deep learning](#) architecture with speaker-dependent features for [speaker identification](#), to ensure real-time operation.

### REVE Systems, Dhaka, Bangladesh

July 2014 - January 2015

*Jr. Software Engineer*

- Programmed controller of ‘media-gateway’ server to facilitate both phone calls and faxes between the [telephone network](#) and [VoIP network](#) or another telephone network. Also designed a front-end panel of controllers for [VoIP administrators](#) and [customers](#).

## TEACHING EXPERIENCE

---

### Indiana University, Bloomington, IN, USA

Fall 2016 - Now

*Teaching Assistant, Research Assistant*

- Mentored 2 undergrad students [Muhammad Asghar](#) (currently works as Software Engineer at [L3Harris](#)) and [Daniel Quintans](#) (currently works as Software Development Engineer Intern at [Amazon](#)) as a part of the [NSF Research Experience for Undergraduates \(REU\)](#) program.
- Mentored **summer visitor undergrad** student [Chitrang Gupta](#) (currently works at the [University of Texas, Austin](#)) in audio focused machine learning research project.
- Currently teaching over 50 undergrads in Data Science course - Data Representation (D-321) and coordinating AIs.
- Taught over 25 undergrads on Computer Science course - Machine Learning (B-455).
- Taught over 60 undergrads on Computer Science course - Python (A-201).
- Taught over 60 undergrads on Computer Science course - Discrete Mathematics (C-241).

### United International University (UIU), Dhaka, Bangladesh

February 2015 - August 2016

*Lecturer, Department of Computer Science & Engineering*

- Evaluated and judged undergrad final research thesis projects.
- Coached undergrad students for **programming contest** at south-asia region of [International Collegiate Programming Contest \(ICPC\)](#), 2016.
- Taught over 2000 undergrads on Computer Science programming courses - Structured Programming Language (CSE-1111), Structured Programming Language Laboratory (CSE-1112), Object Oriented Programming (CSE-1115), and Object Oriented Programming Laboratory (CSE-1116).
- Taught over 800 undergrads on Computer Science hardware courses - Digital Logic Design (CSE-1325) and Digital Logic Design Laboratory (CSE-1326).
- Taught over 60 undergrads on Computer Science algorithm course - Data Structure and Algorithms I (CSE-2215).
- Taught over 250 undergrads on Computer Science artificial intelligence (AI) courses - Artificial Intelligence (CSE-3811) and Artificial Intelligence Laboratory (CSE-3812).

## TALKS & PRESENTATION

---

- Title “Knowledge Distillation on Joint Task End-to-End Speech Translation”, Presenter in **Amazon Services LLC**.
- Title “Unified Speech Enhancement Approach for Speech Degradations and Noise Suppression”, Presenter in **Microsoft Corporation**.
- Title “Incorporating Embedding Vectors from a Human Mean-Opinion Score Prediction Model for Monaural Speech Enhancement”, Presenter in conference **INTERSPEECH 2021**. ([video](#))
- Title “Towards an ASR approach using Acoustic and Language Models for Speech Enhancement”, Presenter **ICASSP 2021**. ([video](#))
- Title “Speech Enhancement”, Presenter in **BOSE Corporation**.
- Title “Monaural speech enhancement using intra-spectral recurrent layers in the magnitude and phase responses”, Presenter in conference **ICASSP 2020**. ([video](#))
- Title “Speech Emotion Recognition”, Guest Lecturer in course **CSCI-B 659**.
- Title “Machine Learning & Deep Learning”, Guest Lecturer in [NSF Research Experience for Undergraduates \(REU\)](#) program, 2019.
- Title “Improved Speech Enhancement by Incorporating Intra-Spectral Dependencies”, Presenter in workshop [Midwest Music and Audio Day \(MMAD\)](#), 2019.
- Title “Incorporating Intra-Spectral Dependencies with a Recurrent Output Layer for Improved Speech Enhancement”, Presenter in conference **MLSP 2019**.

## AWARDS & SERVICES

---

- **Reviewer** in the [International Speech Communication Association \(ISCA\)](#), INTERSPEECH 2023, Dublin, Ireland.
- **Student Volunteer** in managing the virtual platform at the [ISCA](#), INTERSPEECH 2021, Brno, Czech Republic.
- Received **Travel grant** from Indiana University to attend at the IEEE [ICASSP](#) 2020, Barcelona, Spain.
- Received **Travel grant** to attend at the Annual Conference of the [ISCA](#), INTERSPEECH 2019, Graz, Austria.
- **Student Volunteer** for organizing conference at the Annual Conference of the [ISCA](#), INTERSPEECH 2019, Graz, Austria.
- **President** of [Bangladesh Student Association \(BDSA\)](#), Indiana University, Bloomington, IN; which involves actively managing the overall student organization of Bangladeshi students of 50+ members.
- **Vice-President, General Secretary, Treasurer** of [BDSA](#), Indiana University, Bloomington, IN over consecutive years from 2017 to 2019.
- **Panelist**, at career development of NSF [Research Experience for Undergraduates \(REU\)](#) program, 2019.
- Received **Brown belt** in Taekwondo and judged **Yellow belt** students at [Indiana University Taekwondo club](#) from 2017 to 2020.