Soumya Dutta

(+91)8828290596
soumyadutta@iisc.ac.in
https://soumya-dutta.github.io

Research Interests

- Speech Processing
- Multimodal Emotion Recognition
- Synthesis of Emotional Speech
- Deep Learning Approaches for Emotion Recognition and Synthesis

Education

2021-Current **Indian Institute of Science Bangalore**.
Ph.D in Electrical Engineering **CPI: 10/10**

2018 **Indian Institute of Technology Bombay**.
M.Tech in Control and Computing **CPI: 9.41/10**

2015 **Indian Institute of Engineering Science and Technology Shibpur**.
B.E. in Electrical Engineering **CPI: 8.99/10**

Work Experience

Jan 2024-July 2024 **Position: PhD Research Intern, Samsung Research Institute Bangalore**.
Team: Speech and Language Technologies

- Worked on the problem of **textless speech-to-speech translation** system for French-English, Spanish-English and German-English language pairs
- Developed a baseline system inspired by an existing work called TranSentence for this task
- Implemented **expressive speech translation** on top of the baseline system

Jul 2018-Feb 2021 **Position: Cognitive Data Scientist**.
Organization: **IBM**

- Worked on a **SVM Classifier** for email intent classification with a **precision of 76%** and **recall of 91%**
- Worked on a **Virtual Makeup Try-On** system with lips and hair segmentation followed by color transfer from example lipstick and hair-dye patches to lips and hair respectively. The color transfer was done by **matching the distribution** of the source and target.
- Detected human beings in a video by using a **YOLOv5m network pre-trained on COCO dataset**. This was used along with a **tracking algorithm (SORT)** to raise an alert if a person was loitering in an area

Publications

Journals

- **S.Dutta**, S.Ganapathy, "**Leveraging Content and Acoustic Representations for Speech Emotion Recognition**" IEEE/ACM Transactions on Audio, Speech, and Language Processing (*Under Review*)

Conferences

- **S.Dutta**, S.Ganapathy, "**LLM supervised Pre-training for Multimodal Emotion Recognition in Conversations**" ICASSP 2025
- **S.Dutta**, S.Ganapathy, "**Zero Shot Audio to Audio Emotion Transfer With Speaker Disentanglement**" ICASSP 2024
- **S.Dutta**, S.Ganapathy, "**Multimodal Transformer With Learnable Frontend And Self Attention For Emotion Recognition**", ICASSP 2022
- **S.Dutta**, N.Rangaraj, M.N.Belur, S.Dangayach and K.N.Singh, "**Construction of periodic timetables on a suburban rail network-case study from Mumbai**", RailLille 2017

Scholastic Achievements

- **Qualcomm Innovation Fellowship 2023 Finalist** (awarded to 37 out of 124 teams)
- Selected to participate in the **Google Research Week** 2023 organised by Google Research India
- Recipient of the **Prime Minister's Research Fellowship(PMRF)**, 2022
- Awarded the **SPS travel grant** for attending ICASSP 2022 (awarded to 56 out of 3967 papers)
- Recipient of the **MHRD, Government of India** Scholarship for Graduate Studies
- Secured a rank of **231** in **WBJEE 2011** out of approximately *1,30,000 students*

Relevant Course Work

Ph.D	Machine Learning for Signal Processing, Advanced Deep Learning, Stochastic Models and Applications, Optimization for Machine Learning and Data Science, Detection and Estimation Theory, Speech Information Processing
Graduate	Machine Learning, Applied Linear Algebra, Statistical Signal Analysis, Optimization Models

Academic Projects

May-June 2021 **Multimodal Emotion Recognition**.
Advisor: **Dr. Sriram Ganapathy**
Course Name: **Machine Learning for Signal Processing** :

- Trained a **Transformer** model to detect sentiment from videos of IEMOCAP database using speech and provided text transcriptions
- The text features were extracted from a **BERT** based sentiment classifier
- The **accuracy** achieved was 77.8%

Oct - Nov 2021 **Explainability in Audio Classification**.
Advisor: **Dr. Sriram Ganapathy**
Course Name: **Advanced Deep Learning** :

- Implemented a simple classifier on AudioMNIST data and used Layer Relevance Propagation technique for explaining the classifier outputs
- Added distractors to each audio sample to find out the real performance of the explainability of the network

Nov-Dec 2021 **PowerSGD for Efficient Gradient Compression in Distributed Optimization**.
Advisor: **Dr. Sundeep Chepuri**
Course Name: **Optimization for Machine Learning and Data Science** :

- Implemented the PowerSGD algorithm using Pytorch for more efficient distributed optimization in deep learning problems
- The method was tested for two problems of Image classification and Text Sentiment Analysis using BERT

Dec 2019-Feb 2020 **Google QUEST QA Labeling**.
Kaggle

- A **BERT** model for predicting scores for **30 classes based on QA pair** was trained
- Achieved a rank of **72 out of 1571 teams** with a **Spearman's Correlation Coefficient score of 0.39884**

Skills

Programming	Python, C, C++, Pytorch
Software	MATLAB, LaTeX, MS-Office