

Muhammad Umar Farooq, Ph.D.,

✉ umar@emotech.co

mufarooq40@gmail.com

in [/mufarooq40](#)

🌐 umar17.github.io

Professional Experience

- **Applied Research Scientist** July 2024 - Present
Emotech Ltd. *United Kingdom*
 - Conduct applied research in speech-to-text technology and collaborated closely with the engineering team to develop and deploy production-ready solutions
- **Doctoral researcher** Feb 2021 - July 2024
LivePerson Centre for SLT, Dept. of Computer Science *University of Sheffield, UK.*
 - Conducted research on multilingual speech recognition **funded by LivePerson Inc.**
 - Investigated cross-lingual similarities [6], and presented novel multilingual and cross-lingual model fusion techniques [5]
 - Proposed an innovative data augmentation method [3] and bridged a knowledge distillation gap for cross-lingual model adaptation [2]
 - Developed progressive unsupervised domain adaptation and generalisation methods [1] for self-supervised speech recognition models [4]
- **Pre-doc Intern** Sep 2020 - Dec 2020
Institute of Formal and Applied Linguistics (UFAL) *Charles University, Prague.*
 - Funded by **European Union (EU)**
 - Adapted language models for ASR in the **EU Horizon 2020** project European Live Translator ([ELITR](#))
- **Research Assistant** Jun 2017 - Aug 2020
Center for Language Engineering (CLE) *University of Engineering and Technology Lahore*
 - Designed a corpus, developed software for semi-supervised collection [7] and curated Urdu speech data set in collaboration with linguists
 - Built the first Urdu Large Vocabulary Continuous Speech Recognition (LVCSR) system [7] and [deployed it online](#) for users
 - Created an [API for developers](#) as well
 - Adapted Urdu ASR for Urdu-English code-switched spontaneous speech (*IEEE O-COCOSDA, 2020*)
 - Adapted Urdu ASR to Punjabi in a zero-shot approach without using Punjabi speech data
 - Designed and implemented a real-time two-way Urdu-English Speech-to-Speech (S2S) translation system

Education

- **Ph.D. Computer Science** Feb 2021 - May 2024
University of Sheffield *Sheffield, UK.*
 - **Supervisor:** Prof. Thomas Hain
 - **Industry partner:** LivePerson Inc.
 - **Research areas:** Artificial Intelligence (AI), Machine Learning (ML), speech and language processing, automatic speech recognition, foundational models, multilingualism
 - **Thesis title:** Non-linear cross-lingual mappings for low-resource speech recognition
- **M.Sc. Electrical Engineering** Sep 2017 - Jun 2019
University of Engineering & Technology *Lahore, Pakistan.*
 - **CGPA:** 3.68/4.00
 - **Research areas:** AI, ML, speech processing, speaker diarisation
- **B.Sc. Electrical Engineering** Oct 2013 - May 2017
University of Engineering & Technology *Lahore, Pakistan.*
 - **CGPA:** 3.42/4.00

Selected Publications

For full list, please visit the [Google Scholar profile](#) or my [publications web page](#)

- [1] **Progressive Unsupervised Domain Adaptation for ASR Using Ensemble Models and Multi-stage Training**
R. Ahmad, **M. U. Farooq** and T. Hain
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024
- [2] **MUST: A Multilingual Student-Teacher Learning Approach for Low-Resource Speech Recognition**
M. U. Farooq, R. Ahmad and T. Hain
IEEE Automatic Speech Recognition and Understanding (ASRU), 2023
- [3] **Learning Cross-lingual Mappings for Data Augmentation to Improve Low-Resource Speech Recognition**
M. U. Farooq and T. Hain
Interspeech, 2023
- [4] **Towards Domain Generalisation in ASR with Elitist Sampling and Ensemble Knowledge Distillation**
R. Ahmad, M. A. Jalal, **M. U. Farooq**, A. Ollerenshaw and T. Hain
IEEE ICASSP, 2023
- [5] **Non-Linear Pairwise Language Mappings for Low-Resource Multilingual Acoustic Model Fusion**
M. U. Farooq, D. A. H. Narayana and T. Hain
Interspeech, 2022
- [6] **Investigating the Impact of Cross-lingual Acoustic-Phonetic Similarities on Multilingual Speech Recognition**
M. U. Farooq and T. Hain
Interspeech, 2022
- [7] **Improving Large Vocabulary Urdu Speech Recognition System using Deep Neural Networks**
M. U. Farooq, F. Adeeba, S. Rauf and S. Hussain
Interspeech, 2019

Teaching and Mentoring Experience

- **Teaching Assistant** Feb 2024 - June 2024
University of Sheffield, UK.
 - Developed a short course '[Basics of AI for Social Science: Unsupervised Machine Learning](#)' for non-experts (faculty members and PhD students of social sciences)
 - Delivered a short course '[Python Programming](#)' on Python programming language to non-experts
- **Graduate Teaching Assistant (GTA)** Sep 2022 - Feb 2024
University of Sheffield, UK.
 - Assisted module leaders in delivering the lab sessions to the undergraduate and postgraduate students
 - Marked the assignments and exams
 - **Modules:** COM6018, COM6502, ... [Full modules list here.](#)
- **Mentoring / Supervising Interns**
LivePerson Centre for SLT, University of Sheffield
 - Mentored D. A. H. Narayana for his short project on low-resource speech recognition at LivePerson Centre for SLT

Conference organisation and relevant experience

- **UK Speech 2023:** Sheffield, UK.
IT Support Chair, Local Organising Committee UK Speech 2023
- **ISCA-SAC student events** at Interspeech 2023
Member Organising Committee, ISCA Student Advisory Committee (SAC) 2023

Skillset

Technical Skills

- **Languages:** Python, Rust, Java, Shell scripting, L^AT_EX, MATLAB, C
- **Frameworks:** PyTorch, HuggingFace
- **Libraries:** Scikit-learn, NumPy, Pandas, Matplotlib
- **Computing:** SGE/SLURM for remote grid computing
- **Tools:** SpeechBrain, Kaldi, ...
- **Environments:** Linux, MacOS, Windows

Soft Skills

- Problem Solving
- Self-motivation
- Adaptability
- Effective communication
- Technical writing
- Team collaboration
- Time management

Language Proficiency

- English (professional proficiency)
- Urdu (native)
- Punjabi (native)
- Hindi (spoken proficiency)
- Arabic (beginner)

Workshops & Courses Attended

- **International Spring School of Advance Language Engineering (ISSALE 2019)** Moratuwa, Sri Lanka.

A two-week spring school by **DAAD** covering the following courses;

- Deep Learning in Natural Language Processing (NLP)
- Grammar Engineering for South Asian Languages
- Grammar Foundations
- Computational Linguistics

Achievements, Awards and Honours

- **Postgraduate Publication Scholarship, University of Sheffield** Mar 2024 - May 2024
- Submitted my **PhD in 3 years**
- Published **6 research papers** during 3 years of my PhD (4 as a first author and 2 collaborated works)
- **Departmental Scholarship Award, University of Sheffield** (*Fully funded Ph.D.*) Feb 2021 - Jan 2024
- **Excellence Award for Research and Development** for 3 consecutive years at CLE (2017 - 2019)

Memberships

- **International Speech Communication Association (ISCA)** 2022 - 2024
Student member
- **ISCA Student Advisory Committee (ISCA-SAC)** 2023 - 2024
Volunteer
- **Institute of Electrical and Electronics Engineers (IEEE)** 2023
Student member