Statement of Purpose

Omar Fathy · MSc Speech and Language Processing, University of Edinburgh

◆ ◆ ◆

In my graduation project, a speech recogniser that my team fine-tuned on Egyptian Arabic got 29% of words wrong. On sentences where the speaker switched to English halfway through, it got more than 40% wrong. That gap is the problem I want to study: how speech recognition can treat the way many Arabic speakers actually talk, moving between dialect and English in one breath, as normal input rather than noise.

The project began with a practical complaint. Voice search on our phones handled formal Arabic tolerably and Egyptian Arabic badly. With six classmates I recorded and transcribed 30 hours of everyday Egyptian Arabic, mostly conversations about shopping, transport and health, and we fine-tuned an open speech model on it. Word error rate fell from 48% to 29%. The code-switched sentences barely improved, and when I looked at them closely, the errors clustered at the switch points: English words spelled out as Arabic nonsense, Arabic words forced into English spellings.

I graduated from Cairo University in 2024 with a BSc in Computer Science and a GPA of 3.6 out of 4. For the past year I have worked as a machine learning engineer at a start-up in Cairo that builds voice ordering for pharmacies, where I train and evaluate models in Python and PyTorch and have learned the basics of Kaldi. The job has taught me how much evaluation matters: a model that looks good on a clean test set can fail on a real customer in a noisy shop. I have taken courses in probability and linear algebra, and I want the deeper grounding in phonetics and statistical modelling that so far I have picked up only from papers.

At work I have since run a smaller version of the same experiment. Adding 200 English drug and product names, recorded by Egyptian staff, to our training data cut errors on those names by a third but did nothing for switching in general: the model learned the words, not the pattern. Collecting data has taught me something else as well. Every speaker in our university project signed a consent form, and we removed names and phone numbers from every transcript before training. I want to work in a field that treats the people whose voices make up the data with that kind of care.

The MSc in Speech and Language Processing at Edinburgh suits this plan because it treats speech as both a linguistic and a computational problem. Its courses in speech processing and machine learning for language would give me the foundations I lack, and the School of Informatics has a long record in speech technology, so I would be learning among people who have worked on these questions for decades. I hope to write my dissertation on code-switched recognition, starting with a simple test: whether models do better when they are told where a switch is likely to happen.

After the degree I want to build speech tools for people who mix languages, either at a company that serves Arabic-speaking customers or in a research group. In both places the test is the same: whether a pharmacist in Cairo can take an order by voice without asking the customer to speak like a newsreader.

Written with HumanProse