Overcoming Challenges

Three years ago I decided I was going to write a PhD on speech recognition and deep learning.

This decision posed many challenges. No one in my linguistics department works on speech recognition, and I didn't know much about it to begin with. What's worse, my math background was rusty and my machine learning was non-existant. In a nutshell, I had a lot to learn and no one to teach me. Thankfully, support from my professors was never an issue. I began my journey sailing off into unknown territory, knowing my professors were always rooting for me back on shore.

My approach to learning difficult concepts has been the following: teach myself as much as I can, and go get help when I hit a wall. As a result, I have become a better learner, better communicator, and more confident researcher. I have learned to formulate my questions well, listen to what is being said, and most importantly, I'm not afraid to fail. The fear of failure, the fear of looking stupid, held me back from seeking help for the first year. That year was marked with caffiene fueled late nights in the lab, teaching myself linear algebra, probability theory, and backprop. My progress was slow, so I worked harder. Asking for help didn't seem like an option. However, it soon became obvious that I couldn't learn everything from books, papers, and blog posts alone. I needed answers to my questions. I needed intuitions. I needed to learn from the experts. The only problem was, I had to go find them.

If the mountain won't come to Mohammed...

My dissertation topic is multi-task learning for deep neural net acoustic modeling in speech recognition for low-resource languages. Each slice of this topic is substantially different. Mutli-task learning is dominated by NLP or computer vision literature, whereas deep neural nets are pretty much everywhere. However, I decided that I needed help from experts in the area I was hitting the most walls: acoustic modeling for low-resource languages. So, I found the experts, and decided to reach out. Jean-Luc Gauvain and Lori Lamel lead a research team at CNRS near Paris, specilizing on speech recognition for low-resource languages. I reached out, asking if I could learn from them as a visiting scholar. They said yes. Together, we applied for an NSF grant and funding from the French Embassy to the US, and were lucky enough to get both. Along with my NSF graduate research fellowship, I had enough funds to work out of their lab for a year. So I packed by bags and headed for Paris.

However, even after I arrived I was scared to ask questions. Lori got her PhD from MIT in Electrical Engineering, and Jean-Luc wrote *the* paper on MAP adaptation for GMMs. I felt out of my league. However, there came a day where I had a question that I couldn't figure out, and I knew the people in the next room had the answer. It was a very fundamental question about Expectation-Maximization of Hidden Markov Models. I understood the algorithm after it got started (it's an iterative procedure), but I couldn't wrap my head around the very

first step: parameter initialization.

I took my time to make sure I could express the problem well, even drawing out a diagram, and then went over to ask Lori and Jean-Luc. Jean-Luc happened to be around, and I showed him the diagram and asked the question. After some back-and-forth clarifications, I had my answer, and no one thought I was stupid for asking! Later that day, at lunch with the post-docs, I mentioned that I was really happy that I asked the question and realized it was no big deal. They reaffirmed that I should ask as many questions as possible, and to my surprise they were also interested to hear Jean-Luc's answer to my question! From that day on, I spent more time talking with the researchers and post-docs, learning more at the lunch table than from my textbooks.

My experience at the lab in Paris taught me a lot about communicating efficiently. I was the token linguist embedded with the electrical engineers, and we learned how to talk, think creatively, and bounce research ideas off of each other. Since then, I've taken every chance I get to learn one-on-one from researchers in speech recognition and machine learning. I spent a week at the University of Edinburgh learning about speech synthesis with neural nets and an afternoon meeting Dan Povey's (creator of the Kaldi speech recognition toolkit) lab group in Baltimore. The visit to Povey's lab laid to rest some questions I'd had about the Weighted Finite State Transducer technology which is core to Kaldi's decoding graph. Over the past few years, I've learned how to identify walls when I hit them, and how to efficiently get the information I need when that happens.

My Research Passion & How Google Can Help

My research passion stems from a problem I see in the world. I believe everyone has the right to access high quality speech technology, no matter what language they speak. This goal is a long way off, but I'm working towards it.

People from wealthy countries have access to life-changing technologies, while people in poorer countries don't. This is not an exaggeration when I say life-changing. During my work with blind people in Kyrgyzstan, I've heard multiple times how the lack of audiobooks in Kyrgyz schools hinders students from going on to university. Even Google Translate still doesn't have a speech synthesizer for Kyrgyz.

This discepency comes from the lack of labeled training data from these languages. I beleive it is possible to train high quality neural nets on small datasets, but currently it's not a priority. Researchers now spend a lot of time working on big datasets, and so they don't worry about overfitting. My research investigates auxiliary task-creation for multi-task learning with minimal domain-expertise. My research investigates the regularization effects of a single neural net learning multiple tasks, where the tasks were either automatically discovered or defined by expert (linguist) knowledge.

Intuitively, I'm working to make better use of small datasets, because collecting massive datasets for every new classifier is not a viable solution to building speech technology for the world's 6,000 languages.

As a Google AI resident, I will have all the experts to help me accomplish

my goals. I want to publish influential work, and get more people caring about low-resource languages. This Google residency is the perfect environment for me to learn from and collaborate with people who are dedicated to their work and will push me to higher acheivement. I see the list of languages in Google Translate without speech recognition or speech synthesis boxes waiting to be checked off, and I want to help make that happen.

Thank you for your time and consideration.