Cover Letter 2.0

A lot has changed since last year. Last year, I thought Mountain View was a city from some dystopia future, swarmed with self-driving cars and coffee-delivering robots. Now, after last year's on-site interview for the Google AI Residency, I know that Mountain View is actually a peaceful little town with good biryani.

Last year I wasn't very keen on moving to California — Brain Montréal was my top preference for the Residency. This year, I'm writing this cover letter from my apartment in San Francisco, where I'm an NSF intern at Mozilla's Machine Learning Research group, working on the DeepSpeech team.

Last year, I felt my research wasn't good enough for the top machine learning conferences. Since then, I presented at Google's 2018 Workshop on Machine Learning in Speech and Language Processing, I sent a first-authored paper to NAACL-HLT, and just a few days ago I sent off a first-authored paper to ICML. More importantly, I am proud of all of these papers.

But enough about last year — I've got to tell you about all the cool research I'm working on! In the following you'll get an overview of who I am, what I do, how Google AI can help me achieve my goals, and how I can be a positive addition to the Google AI team.

My Research & Why It's Important

Everyone should have access to speech technology in their native language.

This is the conviction that drives my research. I firmly believe that research is important insofar as it has a positive impact in people's lives.

When I started my PhD, my heart was set on theoretical linguistics and cognitive science, but after a year of fieldwork — researching phonetics in Kyrgyzstan — I switched my focus to Automatic Speech Recognition. When Kyrgyz-speakers would ask me what my research was about, I'd say "I'm studying the differences among vowels in the Kyrgyz language." However, when they asked me why my research was important *to them*, I never had a good answer. In the end, I never came up with a good answer, so I changed my research.

I knew that speech technology was how I could make a difference. After teaching myself speech recognition from the ground up, I made the first speech recognizer for the Kyrgyz language, and later, the first Kyrgyz speech synthesizer. Now, Kyrgyz text-to-speech (eSpeak-NG) is freely available for use in Mac, Android, Windows, and Linux! However, there is still a lot of work to be done on speech technology for Kyrgyz (and the rest of the world's 7,000 languages!). Three years after adding Kyrgyz to Google Translate, Google still has no Kyrgyz text-to-speech.

I have a special place in my heart for the Kyrgyz language, but I know it isn't unique in it's lack of resources. The vast majority of the world's languages still don't have speech-to-text or text-to-speech. My research aims to fix this problem.

Currently, I work on Transfer Learning for end-to-end speech recognition because it is the fastest path to speech-to-text for every language. End-to-end speech

recognition is ideal because it doesn't require hand-made linguistic resources for each language (i.e. you don't need a linguist to create a pronunciation dictionary for every language). However, all end-to-end methods are data-hungry. Currently, to train an end-to-end speech recognition model, you should have over 10,000 hours of speech which has been transcribed. Even for companies like Google, it's impossible to get that much data for every language in the world. This is where Transfer Learning comes in.

The goal of Transfer Learning is to extract knowledge from one dataset which can help train a model for another dataset. The first dataset is called the "source-domain" and the second dataset is the "target-domain". For example, the "source-domain" might be English speech, and "target-domain" might be Kyrgyz speech.

The intuition as to why Transfer Learning should work is the following: the "source-domain" contains knowledge which is helpful in understanding the "target-domain", and if that useful information can be extracted, then learning a new task in the target-domain should be easier. For example, any model which performs speech-to-text should know that a honking car can never be part of any language. This separation of background noise from human speech is language-independent, and we should be able to transfer this knowledge from a large English dataset to a model trained for Kyrgyz.

My dissertation focuses on Multi-Task Learning, which is a kind of Transfer Learning that teaches one model to perform multiple tasks in parallel, exploiting knowledge from each task. Currently, with the DeepSpeech team at Mozilla, I'm working on another kind of Transfer Learning — model-based transfer. In my most recent work (ICML 2019 submission), my co-authors and I show which layers of a pre-trained English model transfer best to a new language — with results from 12 different languages!. We also use experiments to interpret what exactly the model learned at different layers. There's a lot of exciting work to be done in Transfer Learning in end-to-end speech recognition, and I want to be on the team doing it!

Challenges of Open-Ended Research

I've been doing open-ended research since my time as an undergrad research assistant, when my lab Principal Investigator suggested that I come up with my own research project. Since then, the biggest challenge has been knowing when to ask for help. I've never had a problem coming up with new research ideas, but for a long time I tried to do everything on my own. I now know that without the help of others, I cannot accomplish my research goals.

No one at my university works on speech recognition, and so I decided to get help from experts. I spent a year learning fundamentals of speech recognition for low-resource languages as a visiting student at the LIMSI Research lab at CNRS in Paris. At Mozilla Machine Learning Research, I have the freedom choose what research I work on, but I know that doing it alone is impossible. I've learned how to communicate with my team effectively to first agree on a project that is interesting to everyone and then split up the work so that each person is working on what they do best. For the last two papers, I came up with the main research direction, a colleague (a real TensorFlow guru) did most of the coding, and after

I ran the experiments, all colleagues helped interpret the results. In short, I've learned that in this field you can't do research alone — being a part of a productive and supportive lab is crucial, and I know that is exactly what Google AI & Brain have to offer.

Achieving my Goals with the AI Residency

Quite simply, working with and learning from the world's leaders in machine learning will help me advance my goals because you know how to implement research that scales. I want to create speech-to-text and text-to-speech for every language in the world, and Google has access to tons of data, more compute than I could dream of, and also, a friendly, supportive research environment. I constantly have new research ideas for Transfer Learning, Multi-Task Learning, and end-to-end speech technologies — the researchers at Google will be able to help me decide which ideas are worth following and how to follow the best ideas effectively. In short, the Google AI Residency will allow me to do research that scales, and I know I can be a positive, creative addition to the team.

Thank you for your time and consideration.

Checking off boxes

1. Research Blog: jrmeyer.github.io

2. GitHub: JRMeyer

3. TensorFlow skills: The Flow of TensorFlow & The Joy of the TensorFlow Dataset & Estimator APIs

4. Math skills: MLE for Gaussians tutorial

5. Speech Synthesis tutorial: Getting started with Merlin

6. Multi-Task PhD Research: Multi-Task Kaldi

7. Graduate Coursework: Applied NLP // Statistical NLP // Intro to Machine Learning // Speech Language Technology // Regression Analysis (A + B)