

Cover Letter 2.0

A lot has changed since last year. Last year, before I arrived in Mountain View for the final round of interviews for the Google AI Residency, I thought that Silicon Valley would be swarmed with self-driving cars and coffee-delivering robots. Now, I know that Mountain View is actually a peaceful little town with good biryani.

Also, last year I wasn't very keen on moving to California — Brain Montréal was my top preference. 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 on the DeepSpeech team.

In September, I took a major step in my research when I attended Interspeech in Hyderabad, India. Attending Interspeech has been a dream of mine ever since I started working in speech recognition. When I started my PhD my heart was set on cognitive science and linguistics, but after the 2nd year I switched my focus to automatic speech recognition. I wanted to do more impactful research. After teaching myself speech recognition from the ground up, I finally made it to Interspeech. I presented my work at Google's co-located workshop on Machine Learning in Speech and Language Processing (Meyer 2018) and received lots of great feedback from Googlers and other workshop attendees.

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 Current Research

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

This is the conviction that drives my research. Currently, I work on Transfer Learning for end-to-end speech recognition because I believe it is the fastest path to ASR in every language. End-to-end ASR is ideal because it doesn't require linguistic resources. However, in its current form end-to-end ASR doesn't work for low-resource languages because it requires around 10,000 hours of audio. 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 use source-domain knowledge as an inductive bias while training a model for some target-domain task. For example, the "source-domain" might be English speech, and "target-domain" might be speech from the Kyrgyz language. Source-domain knowledge can be found either *directly* in a dataset, or *indirectly* in a trained model. I've worked with both of these approaches, and I believe some combination of them is the future of the field — especially when it comes to small datasets. With Transfer Learning you could use just a few hours of speech from Kyrgyz in combination with a large English dataset to produce a model that works better than training on Kyrgyz alone. Transfer learning isn't new, but applying it to end-to-end ASR is very new, and very relevant.

My dissertation focuses on Multi-Task Learning, which is a *direct* transfer learning approach. I show that using linguistic categories as labels for classification (in addition to the standard output layer) improves recognition on small datasets. Furthermore, I show that it is possible to discover these linguistic classes without the need of a linguist, making the approach truly scalable.

Currently, with the DeepSpeech team at Mozilla, I'm working on the parameter-copying approach of transfer learning — an *indirect* transfer method. This approach is both simple and elegant. In my most recent work (NAACL 2019 submission), I show that copying layers from a source model transfer best to a new language if the recurrent layers are trained from scratch. However, any preceding fully-connected layers should be copied and frozen straight from English.

There's a lot of exciting work to be done in transferability of end-to-end ASR models, and I want to be the one doing it! For image recognition models we already have some idea as to which features are encoded at which layers, but for speech recognition we have no idea.

Why is it Important?

End-to-end Transfer learning is the most promising approach for making sure that ASR accessible for all languages, ages, and genders. In every case, this is a small-data problem. If we can develop robust methods for training (or adapting) these end-to-end models on small datasets, then we can be sure to cover more languages, more accents, and be less biased towards certain genders (Tatman 2017).

Furthermore, for researchers and developers in developing countries transfer-learning from a source model is much more feasible, because it can be implemented on a single GPU or even a CPU, regardless of how many GPUs or TPUs were used to train the source model. In short, with transfer learning and multi-task learning I want to push for the BERT moment in ASR.

Open-Ended Research

I've been doing open-ended research from the beginning of my PhD, and the biggest challenge has been knowing when to ask for help. No one at my university works on ASR, and so I learned quickly that without guidance it's impossible to make real progress. I spent a year learning the fundamentals as a visiting student at the LIMSI lab at CNRS in Paris, and now at Mozilla I'm learning about the distributed computing and more productive research workflows. Most importantly, 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.

The AI Residency

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)

Thank you for your time and consideration.