

## My Story

Seconds before they started filming, all I could think about was how the lights were too bright. I thought about how my pants rode up on my ankles as I sat on that uncomfortable polyester couch. I repeated over and over in my head how to say “practical application” in Russian, because for some reason I always forget it. Then, before I knew it, they were done filming and it was all over. We had just presented the first open-source Kyrgyz-language speech synthesizer on the nightly news in Bishkek, Kyrgyzstan. I could finally take a breath.

When I think back to that day, I don’t think about the lights, the couch, or Russian grammar. I think about the strange and wonderful series of events which lead up to that day - many from hard work, and many from pure luck.

I first went to Kyrgyzstan in the Fall of 2013 to start a 10-month long research fellowship on the phonetics of the Kyrgyz language. Funny enough, I started my first foray into coding in the coffeeshops of Bishkek. I wrote an experiment in MATLAB which presented speech sounds to participants, and then recorded their responses and reaction times. When I returned to the University of Arizona, I knew I wanted to continue my work with Kyrgyzstan, programming, and phonetics, but I needed to do more socially conscious research. I wanted to do something that would affect peoples’ lives outside of a handful of theoretical linguists. As luck would have it, that’s when I stumbled upon my first NLP course.

That course showed me that research can be both intellectually satisfying and of practical import. I fell in love with the machine learning mindset. All of a sudden, we weren’t trying to lower some  $p$  value to less than .05; in machine learning we were solving real-world problems.

March 30, 2015, I got an email that changed the course of my life forever. After applying three years in a row, I was awarded an NSF Graduate Research Fellowship. I got three years of funding to do whatever research I could think of, but I knew that time would go fast.

So, I quickly finished up my coursework, and searched for my next *purposeful* research project. Since most of my research ideas center around speech technologies (speech recognition and synthesis) I knew that I needed a speech corpus. For the Kyrgyz language there was no such corpus, so I decided to make one.

For three months in the summer of 2016, I lugged my recording equipment all around Kyrgyzstan and convinced people to let me record their conversations. I teamed up with the American University of Central Asia (the same people who gave me the fellowship in 2013). With their sound studio, connections, and on-the-ground support, I managed to record over 100 hours of conversational Kyrgyz.

While that summer was full of amazing experiences, it taught me the hard truth of working with low-resource languages. Collecting data takes too much time. From that point on, my goal in research has been to combat the bigger problem of data scarcity in machine learning. In this vain, my dissertation explores algorithms which leverage multi-task learning on small datasets, in conjunction with expert knowledge and automated regularization for more efficient learning (ie. better generalization). But back to the story.

That summer I spread the word that I had the skills to make speech technology,

---

and I wanted to collaborate with a local organization on a project. One day sitting around the dinner table, an acquaintance told me about a local organization for blind people that wanted a speech synthesizer. Two weeks later we had a working prototype, completely open-source (eSpeak NG), which could be plugged in to an open-source screen reading software (NVDA). After hours of back-and-forth with the blind Kyrgyz speakers, we had something passable, but far from human-like. This first synthesizer was rule-based, but I knew a neural network based system could be much better.

I came across some papers from Google Research on low-resource text-to-speech and contacted one of the main authors, Alexander Gutkin. I said I wanted help to make an open-source synthesizer, so that Kyrgyz blind people could get easier access. He kindly replied that while he agrees the Central Asian languages need better coverage (from Google Translate in particular), they would not be able to work with me on any project which would be open-source. However, Gutkin recommended that I get in touch with the speech team at the University of Edinburgh, and he introduced me to Simon King, the head of the Edinburgh Center for Speech Technology Research.

King introduced me to his center's text-to-speech toolkits. I like writing tutorials on my blog, so in working with King's main toolkit, Merlin, I wrote up a walk-through. King and his team approved of my post, and now Merlin's main GitHub README features a link to my tutorial.

We kept in touch, and I kept working on my main ASR research with an eye on text-to-speech. Then, King was invited to give a tutorial on deep learning in text-to-speech at Interspeech 2017, and he invited me to fly out to Edinburgh to help out with the tutorial. Practically speaking I played the part of 'typical workshop attendee', and they tested their tutorial on me. I spent a couple hours with each person on the team, and they explained their individual sections to me.

By the end of the week, I had a very solid idea of how text-to-speech works with deep learning, and they had a better idea of how to best deliver their information. After we finished work on the tutorial, I worked closely with Oliver Watts (who works on the text-processing frontend), and we managed to train a Kyrgyz neural net synthesizer from an audiobook. I subsequently wrote a tutorial on how to combine Watts' frontend (Ossian) and King's Merlin in order to train a neural synthesizer for a new language from scratch. Just recently, in December of 2017, I was invited to the Moscow Higher School of Economics to give a workshop on Merlin + Ossian, and with the graduate students from computational linguistics, we created a new synthetic voice for the Chuvash language.

The reason I chose to tell this long and winding story is the following: I know that I am a hard worker, I love research, and I thrive on collaboration. Many of these details and more you will find in my CV, but their context is not. My research continually evolves, my projects spur from old projects, and my collaborations spur from past collaborations. As a Facebook AI Research Resident, I will bring this same dedication and collaborative spirit along with me. Moving forward, I will push myself to the next level in machine learning research. The Facebook AI residency will provide me with the mentorships and resources to do just that.

---

**Checking off boxes**

1. Research Blog: [jrmeyer.github.io](http://jrmeyer.github.io)
2. TensorFlow skills: The Flow of TensorFlow
3. Math skills: MLE for Gaussians tutorial
4. Merlin tutorial: Getting started with Merlin
5. Current Multitask DNN Research: Kaldi Mirror
6. Best Email Classifier in Graduate NLP Class: Ensemble Classifier
7. Graduate Coursework: Applied NLP // Statistical NLP // Intro to Machine Learning // Speech Language Technology // Regression Analysis (A + B)
8. GitHub: JRMeyer

*Thank you for your time and consideration.*