

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 while I was sitting on the uncomfortable polyester couch. I repeated 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.

When I think back to that day, I don't think about the lights, my pants, 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 coffeshops 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 decided that while I loved Kyrgyzstan, programming, and phonetics, I needed to do more applied, impactful research. I wanted to do something that would affect peoples' lives outside of a handful of theoretical linguists. Around that same time, I stumbled upon my first NLP course.

That course showed me that research can be both intellectually satisfying and be of practical import. Furthermore, I fell in love with the machine learning mindset. All of a sudden, we weren't praying to lower our p values under .05, we were trying to solve 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 finished up my coursework as fast as I could, and searched for my next *purposeful* research project. Since all my research ideas were centered around speech technologies (speech recognition and synthesis) I knew that I needed a speech corpus. For Kyrgyz 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 them in conversation. I teamed up with the American University of Central Asia (the same people who gave me the fellowship in 2013-2014). With their sound studio, connections, and general support, I managed to record over 100 hours of conversational Kyrgyz.

That summer I spread the word as much as possible that I had the skills to make speech technology, and I wanted to team up with a local organization on a project. One day sitting around the dinner table, a friend of a friend (Kamen Bonov) said to me he knew an 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 many hours of back-and-forth with the people from the blind organization who use NVDA, we had something passable, but far from human-

like. The synthesizer I made is rule-based, but a statistical parametric speech synthesis system would be better.

I read some papers from Google Research on low-resource text-to-speech and decided to contact one of the main authors, Alexander Gutkin. I said I was looking for 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 tech 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 and I talked back and forth for a while, and he 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 walkthrough. King and his team approved of the quality 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 Kyrgyz 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. What's better yet, I worked closely with Oliver Watts (who works on the text-processing frontend), and he helped me 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, and my projects spur from older projects, and my collaborations spur from past collaborations. As a Google AI Resident, I will bring this same dedication and collaborative spirit along with me.

Thank you for your time and consideration.
