

RELEVANT EXPERIENCE

Machine Learning Language Engineer | Papercup | July 2022 - December 2023 | London, UK

- Used [Hugging Face](#) and [PyTorch](#) to finetune an [LLM](#) system that achieved phoneme error rate of 1.2%, beating our seq2seq g2p system to replace Papercup's laboriously hand-crafted multilingual linguistic front end with an LLM. Models trained with [cloud compute](#) and checkpoints stored on [AWS](#).
- Owned data acquisition, model training, and deployment of 5 new Brazilian Portuguese voices to the product in 2023.
- Used hyperparameter tuning, model architectural changes, and training data ablations with [DVC](#) and [PyTorch](#) to successfully improve 85% of pre-2023 Portuguese voices, reducing the gap between TTS and natural speech by 33%.
- Researched language-specific requirements for TTS in French, Castilian Spanish, US English, Japanese, Hindi, Arabic, and Mandarin to enable the company to weigh engineering costs to decide which languages to expand.
- Spiked [MFA](#) and [PyKaldi](#), and identified PyKaldi as the best [Kaldi](#) alternative for Papercup's forced alignment systems, reducing the team's reliance on Bash in favour of something Python-based.

Data Engineer | Papercup | September 2019 - July 2022 | London, UK

- Developed [Airflow](#) data processing pipeline to automatically manipulate and generate TTS training data and upload it to our data store, replacing the need for researchers to manually process training data before training models.
- Created [Django](#) SQL database to store all TTS training data, including multiple different types of data: audio, numpy arrays, and text, and the relationships between them and metadata about them, enabling researchers to consistently use the same data between experiments.
- Developed a rapid process to commission and add 4 languages in 4 months to Papercup TTS.

Linguist | Adecco at Google | September 2017 - September 2018 | Mountain View, California, USA

- Using [Thrax](#), added over 30 new languages to Gboard in a multitude of scripts and alphabets, including Latin, Arabic, Tifinagh, Persian, Bengali, and Devanagari using data collected from native speakers.
- Corrected user-reported ASR bugs, such as when the Google assistant was mishearing the band U2 and "YouTube", and deployed to production.
- Created spoken emoji grammars for US and Spain Spanish, enabling Gboard users to dictate emojis.
- Performed quality control of scripts and recordings, both in and out of studio, to launch the first US English male voice to the Google assistant

PUBLICATIONS

- [UK Speech](#) Torresquintero, A., Gómez Ibarrondo, T., Wallis, C. G. R., Hu, V., Leoni, J., Ram Mohan, D. S., Hodari, Z. (2023). Incremental Training Changes to Improve Synthesis Quality. UK Speech 2023, Sheffield, UK. (pp. 109). Link: [Google Drive](#)
- [ICASSP](#) Teh, T. H., Hu, V., Ram Mohan, D. S., Hodari, Z., Wallis, C. G. R., Gómez Ibarrondo, T., Torresquintero, A., Leoni, J., Gales, M., King, S. (2023). Ensemble Prosody Prediction For Expressive Speech Synthesis. In Proc. ICASSP 2023. doi: [10.1109/ICASSP49357.2023.10096962](#)
- [Interspeech](#) Torresquintero, A., Teh, T. H., Wallis, C. G. R., Staib, M., Ram Mohan, D. S., Hu, V., Foglianti, L., Gao, J., & King, S. (2021). ADEPT: A Dataset for Evaluating Prosody Transfer. In Proc. Interspeech 2021 (pp. 3880-3884). doi: [10.21437/Interspeech.2021-1610](#)
- [Interspeech](#) Mohan, D. S. R., Hu, V., Teh, T. H., Torresquintero, A., Wallis, C. G. R., Staib, M., Foglianti, L., Gao, J., & King, S. (2021). Ctrl-P: Temporal Control of Prosodic Variation for Speech Synthesis. In Proc. Interspeech 2021 (pp. 3875-3879). doi: [10.21437/Interspeech.2021-1583](#)
- [Interspeech](#) Staib, M., Teh, T. H., Torresquintero, A., Mohan, D. S. R., Foglianti, L., Lenain, R., & Gao, J. (2020). Phonological Features for 0-Shot Multilingual Speech Synthesis. In Proc. Interspeech 2020 (pp. 2942-2946). doi: [10.21437/Interspeech.2020-1821](#)

EDUCATION

Master of Science in Speech and Language Processing | University of Edinburgh | 2018 - 2019 | Edinburgh, UK

76% average (Distinction) in courses. Selected coursework:

- Compared monophone and triphone HMM-GMM models and HMM-DNN models for ASR using Kaldi.
- Built a unit selection speech synthesiser using Festival, with recordings from own voice.
- Implemented multiple methods to compute similarity between sparse word vectors from a Twitter Corpus.
- Built a diphonic speech synthesiser in Python from scratch.

70% (Distinction) on dissertation

- Manually created (from scratch) an implementation of the Wavenet audio generative model in PyTorch.

Bachelor of Arts in Linguistics (Computational) | Yale University | 2012 - 2016 | New Haven, Connecticut, USA

Courses in phonetics, phonology, computational linguistics, syntax, morphology, and computer science. Senior thesis research on stress in American English.

Scholarship awards:

- Sterling Scholar
- M. Albert Geib Scholar

SKILLS

Programming Languages

- Python (7 years)
- Unix/Bash (3 years)
- Thrax (1 year)
- Praat (2 years)
- HTML/CSS (10 years)

Tools and packages: Git/Github, Hugging Face, Kaldi, PyTorch, Festival, PyKaldi, Poetry, Pipenv, Conda

Natural Languages: Spanish (limited proficiency), Arabic - MSA (elementary proficiency), Egyptian Hieroglyphs (elementary proficiency)

BLOG POSTS

2022

- [Overview of TTS at Interspeech 2022](#) (contributor) -- Over 70 likes and 4000 impressions on LinkedIn
- [Highlights of Interspeech 2022](#) (contributor)
- [Papercups go to UK Speech 2022](#) (contributor)

2021

- [Highlights of SSW and INTERSPEECH 2021](#) (contributor)
- [ADEPT: A Dataset for Evaluating Prosody Transfer](#) (lead)
- [ADEPT how-to](#) (lead)

2020

- [Tutorial on declipping wav files](#) (lead)

SUMMARY

I discovered linguistics in my first year at Yale, and instantly fell in love with speech. That something so intuitive and easy for humans to use and understand could be equally as difficult for computers to process enthralled me. This discrepancy only reinforced itself when I was on the Google Assistant team. WaveNet was deployed to production alongside the first US-english male voice, which I helped record the training data for, but no one on my team could explain to me how WaveNet worked. Frustrated by this, I applied (and got accepted to!) Edinburgh's Speech and Language Processing programme, where I not only learned exactly how WaveNet worked, but actually got to manually construct it (from scratch) for my masters dissertation. But I couldn't get enough of speech synthesis, and subsequently started a data engineering role at Papercup, a small 15-person AI dubbing startup, proclaiming that I'd take any job as long as I got to do something with speech synthesis. As Papercup grew, so did my interests. After almost three years I transitioned to a machine learning language engineer role, getting to do more experimentation and research, and owning the training and deployment of new voices to the product. After over 4 years at Papercup, we had almost 70 employees, and even more custom deployed voices, and it was time for me to take a long-needed break. But I'm eagerly looking forward to stretching my machine learning and research brain in my next opportunity.

OTHER INTERESTS

Drag: I've been an active drag king on the London scene since 2022, when I made it to the final of Europe's largest drag king competition, Man Up. London has a centuries-long history of drag kings, with "male impersonators" being one of the most lucrative forms of income for London women in the late 19th century.

Jigsaw Puzzling: I've been an avid jigsaw puzzler for years, and have a particular appreciation for 3D puzzles, as I have an 1800 piece 3D castle at home. I was also a member of Google's competitive puzzling team Silkie Snippets.