Comparison and development of Machine Learning text-to-speech (TTS) system (30HP)

Wiraya is one of Swedens fastest growing startups and offers a Managed Mobile Customer Activation software that inspires action from your best customers. We generate the right messages, to the right people, through the mobile channels they trust, at exactly the right time. Automatically.
We blend voice, text and mobile messaging adapted to each individual and use data and artificial intelligence to (automatically) optimise when, what and how to communicate with each individual.
Since our start we’ve done over 100 millions of communications to customers.

One of our most popular channels, our voice communication, is created in a studio with professional voice actors. While great for quality and user experience it limits our scalability. With the rise of machine learning and intelligent assistants, we’re seeing that speech synthesis have rapidly improved. We’re now interested in how Wiraya can leverage these advances, is it feasible to develop our own system or buy/use off the shelf?

In this thesis work we’re looking for students who have a computational background and an interest in machine learning.

Scope:

Compare current state-of-the-art TTS systems. Special interest in:
Quality (Both subjective and objective metrics)
How well it scales (Both from technical point of view and economical)
Customisability (Different voices, gender, tone of voice)

Develop a prototype TTS system and benchmark against comparable systems above and a recommendation for building ourselves vs buying off the shelf.

Location: Flexible, perferably a few days a week at Wirayas office, Hötorget, Stockholm.

Contact information:
Erik Kurin
erik@wiraya.com

https://deepmind.com/blog/wavenet-launches-google-assistant/
https://machinelearning.apple.com/2017/08/06/siri-voices.html