



Speech Breathing Synthesis for Empathic Virtual Agents

The weekly updates presentation.

Nicolò Loddo



Week 2 - Speech Breathing Empathy Project

Nicolò Loddo

What happened

- I started experimenting with Flowtron. I installed it and used the inference module to synthesize simple text to speech.
- Resolved problems with WaveGlow (mel-spectrogram to audio) to achieve inference.
- Started looking into Style Transfer for Flowtron through a colab notebook that explains how to. It didn't go through for now because of an update of PyTorch. It seems quite easy to solve accessing Flowtron's code
- Found a library that features various TTS models that I have seen in the literature in a convenient framework: <https://github.com/coqui-ai/TTS> and that features a simple model implementation framework
- Read further into spontaneous speech synthesis on the KTH Speech Synthesis demo page which features various studies with interesting results (<https://www.speech.kth.se/tts-demos/>)

- Started some sample observation of the INTERSPEECH dataset
- Extended my literature readings with other examples of spontaneous speech applied to VAs

Next week to-dos

- Try other models on coqui.ai page
- Do a systematic exploration of the INTERSPEECH dataset
- Try to set up the training of a model, maybe using coqui.ai and/or with Flowtron itself, using the INTERSPEECH dataset
- Look more into training spontaneous speech models papers
- Extend literature review

Examples



01

Spontaneous1

<https://www.speech.kth.se/tts-demos/inter-speech2022/>



02

Spontaneous2

https://hfkm.github.io/pc_nhmm_tts/



03

Spontaneous3

<https://www.speech.kth.se/tts-demos/LREC22/>

04

Flowtron

<https://nv-adlr.github.io/Flowtron>



Flowtron

05

Microsoft

<https://azure.microsoft.com/en-us/products/cognitive-services/text-to-speech/#features>