

Week 3 - Speech Breathing Empathy Project

Nicolò Loddo

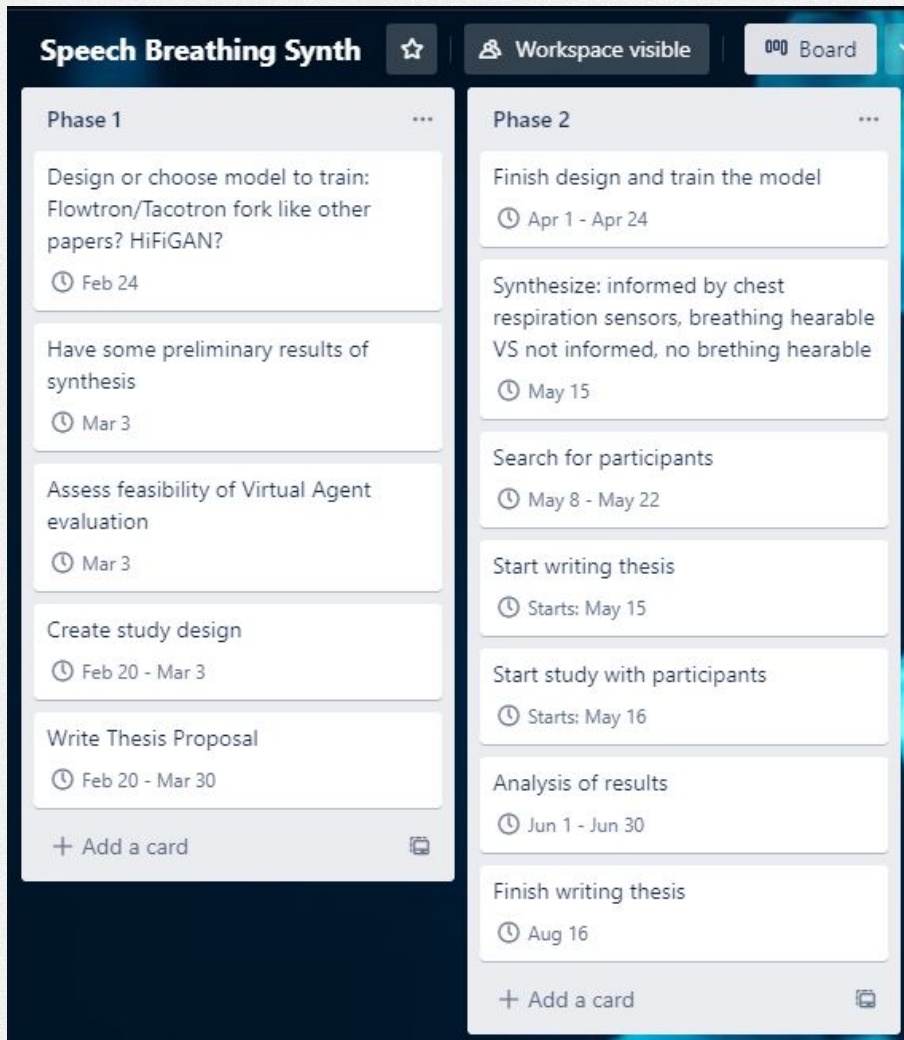
What happened

- Fixed the Style Transfer for Flowtron and applied it with the RAVDNESS Emotional dataset.
- Small exploration of the INTERSPEECH to understand how to exploit the data
- Wrote scripts for transcription of audio files through IBM Watson Speech to Text recognition (API)
- Wrote scripts to use the Gentle Aligner: insertion of Filled Pauses, accurate time alignment and phonemes recognition
- Applied the above said scripts on the INTERSPEECH dataset: I managed to transcribe it all, with filled pauses and phonemes, and to have it aligned with timestamps per phoneme.
- I qualitatively checked consistency of the timestamps with the breathing timestamps, it seems to check out

- Tried Flowtron Style Transfer with the transcribed INTERSPEECH: memory problems due to the length of the audios
- Made schedule
- Found other interesting spontaneous speech models, in particular AdaSpeech seems really good

Next week to-dos

- Search for emotional dataset in english
- Try other models, maybe on coqui.ai
- Adapt a model design for the training of our data. The breath levels are what makes our data special
- Try to set up the training of the forked model
- Read better the newly found spontaneous speech papers
- Extend literature review



HiFiGAN Demo Audio:

<https://jik876.github.io/hifi-gan-demo/>