Multilingual Text to Speech

Software Project-2021/2022

Rasul, Anna, Claesia, Sharmila

Outline

- Project Idea
- Motivation
- Related Works
- Contribution
- Timeline
- Sources

Project Idea:
Multilingual Multispeaker Expressive Text-to-Speech

A single model that takes text in multiple language (10 languages most probably) and converts it into expressive speech.

- ERISHA library: https://github.com/ajinkyakulkarni14/ERISHA
- Our GithHub: https://github.com/sarmilaupadhyaya/Software-Project-2021-UL

Motivation

- Delving into Speech preprocessing and its intricacies.
- Working on the expressivity of speech, given the data is not.
- Multilingual Setting.
- Working as a team to get an End Product that we could be proud of.

Related Work

- Electromechanical - Bell Labs, 1930s
- Rule-based computer synthesis from 1950s and 60s (Klatt, 1987)
- Text→ Norm -> Phonemes -> Phonetics ->Vocoder-> Audio
- Present dominance of end-to-end synthesis (Jurafsky & Martin, 2021)
    - Raw text -> norm -> spectrogram -> Audio
    - Wavenet, Tacotron, Tacotron2

Related Work

- Expressive end-to–end TTS (Kulkarni et al, 2021)
  - Extending tacotron2
  - Speaker encoder
  - Expressivity encoder
- Monolingual French model
  - Neutral and expressive corpora
  - Mean Opinion Score

Contribution

- Model enhancement and training
- Comparison with the previous work
- Evaluation
- Web app creation

Technical Details

- TensorFlow / PyTorch
- Flask + JavaScript for the app

Timeline

28 Sep - project approval

Estimated Time

- Data Preprocessing and Pipeline: 1 week - October 15
- Model Training: 2 weeks - November 7
- Evaluation: 1 week - November 15
- Inference: 1 week - November 30
- UI setup: 2 weeks - October, November
- Integration: 1 week - December
- Report writing: 3 weeks - January

15 Feb - project defense

Sources

- Ajinkya Kulkarni, Vincent Colotte, Denis Jouvet. Improving transfer of expressivity for end-to-end multispeaker text-to-speech synthesis. EUSIPCO 2021, EURASIP, Aug 2021, Dublin, Ireland. ffhal02978485v2f
- Daniel Jurafsky & James H. Martin. Speech and Language Processing.. Draft of September 21, 2021.
- DH Klatt. Review of text-to-speech conversion for English. J Acoust Soc Am. 1987 Sep;82(3):737-93. doi: 10.1121/1.395275. PMID: 2958525.

Thank you ! Any Questions?