Software Development Project

Objective

- Main goal: a web interface for TTS
- Multilingual TTS - English and French
- Input: text
- Output: speech

Datasets

English: LJS - female voice (already preprocessed)

https://keithito.com/LJ-Speech-Dataset/

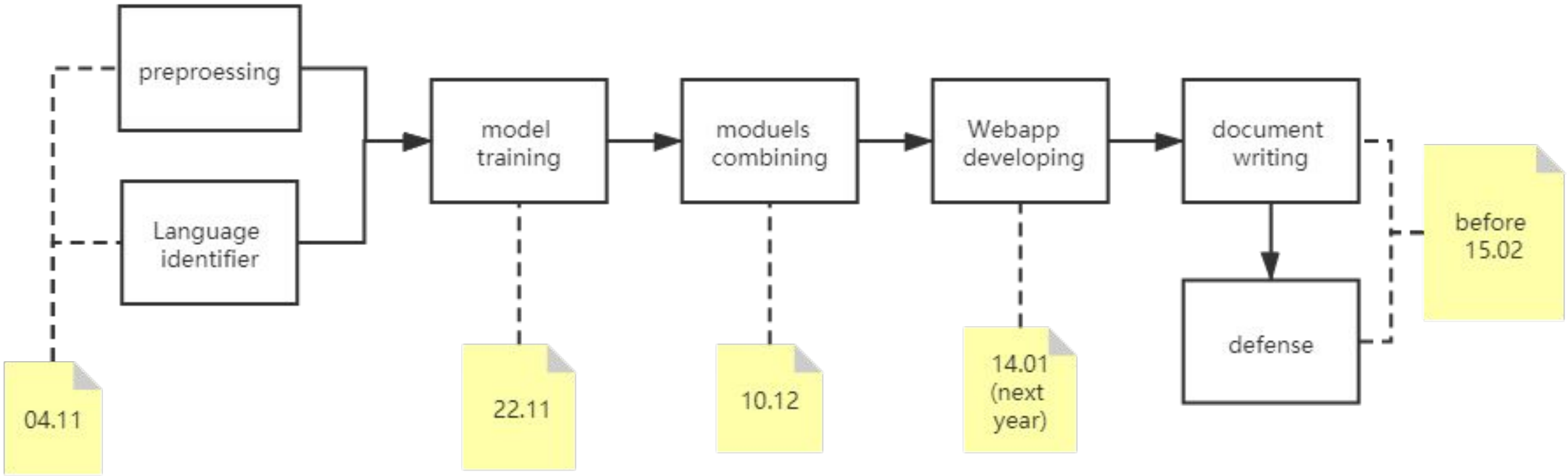
- 13100 utterances (total length ~24h)

French: SIWIS - female voice (**to be preprocessed**)

https://datashare.ed.ac.uk/handle/10283/2353

- 9750 utterances (~10+h)

Work planning



Model : Grad TTS

GradTTS to syntesize the speech from the preprocessed data https://github.com/huawei-noah/Speech-Backbones/tree/main/Grad-TTS

- should already be pre-trained on English
- need to train with French

Given a sequence of phonemes, generate acoustic features as a Mel spectrogram that can be turned into speech waveform

Strategy

- Access to Grid 5k
- Work in pairs first and split when possible
- Read paper (Grad-TTS)
- Possibility of collaboration with the other group for pre-processing, deployment, …
- PyTorch

Current Work

- A repo has been created
- Preprocessing for French

  (1) perl script

  (2) generate the labels & extract main phonemes

- Language Identifier

  (1) LSTM classifier

  (2) French & English monolingual data