

Alena YAKAVETS
Cécile MACAIRE
Chanoudom PRACH
Ludivine ROBERT

Software Project

MULTILINGUAL TEXT-TO-SPEECH SYSTEM

M2 NLP 2020-2021

October 27, 2020



Presentation Overview

- Quick project recap
- Done
- In progress
- To do
- Timeline



Project Recap

- Multilingual TTS, based on Tacotron
 - Tacotron: end-to-end generative TTS model based on the sequence-to-sequence with attention paradigm
 - Multilingual: able to generate speech according to a selected language model
 - Speech generation + emotional contour transfer
- Language/Corpus:
 - English: EmoV-DB, [LJSpeech \(new!\)](#)
 - French: SIWIS



Items Done

- Corpora pre-processing
 - New English corpus: graphemes > phonemes > phoneme ID
 - Metadata file created – ready for model training!

```
/bea/Amused/amused_169-196_0192.wav|9, 7, 0, 18, 0, 2, 26, 3, 19, 13, 35, 18, 2|1|1|0  
/bea/Amused/amused_57-84_0074.wav|15, 2, 0, 2, 11, 0, 11, 2, 36, 14, 19, 8, 16, 7, 2, 22, 13, 24, 22, 10, 3, 18, 29, 20,  
23, 9, 18, 17|1|1|0  
/bea/Amused/amused_1-15_0012.wav|27, 35, 13, 30, 0, 33, 15, 32, 9, 18, 24, 28, 14, 19, 15, 2, 0, 19, 1, 24, 3, 5, 9, 18,  
24, 20, 7, 3|1|1|0  
/bea/Amused/amused_85-112_0091.wav|9, 18, 24, 33, 16, 13, 0, 12, 19, 5, 9, 26, 3, 22, 18, 31, 19, 10, 37, 15, 3, 36, 13,  
31, 21, 3|1|1|0  
/bea/Amused/amused_169-196_0191.wav|20, 23, 13, 3, 8, 17, 36, 7, 18, 32, 6, 8, 3, 18, 29, 31, 19, 18, 16, 13, 30, 9, 18,  
24, 13, 0, 30, 14, 2, 28, 23, 22, 13, 24, 12, 21, 2|1|1|0  
/bea/Amused/amused_16-28_0026.wav|18, 3, 13, 32, 23, 0, 3, 18, 17, 7, 36, 15, 3, 36, 14, 19, 22, 34, 0, 9, 15, 30, 3, 18,  
33, 7, 13, 2, 13, 32, 11, 2, 3, 18, 29|1|1|0
```

<filepath wav>|<phonemesid>|<speakerid>|<emotions>|<languageid>

- Subjective evaluation tests of French speech samples, produced by the base model (thanks for your participation!)



Items Done

Shamcey

☰

NAVIGATION

🏠 List Evaluation

🏠 List Evaluation 2

🔊 Audio

Evaluation / List

No	First Name
1	chanoudor
2	chanoudor

Prev 1 Next

Evaluation Audio

No	Audio	Speaker	Emotion
1	<div>neut_parl_s01_0003.wav</div> <div>▶ 0:00 / 0:01 <div></div> 🔊 ⋮</div>	2	2
2	<div>amused_1-15_0005.wav</div> <div>▶ 0:00 / 0:00 <div></div> 🔊</div>	3	1
3	<div>neut_parl_s01_0002.wav</div> <div>▶ 0:00 / 0:02 <div></div> 🔊 ⋮</div>	3	5
4	<div>neut_parl_s01_0001.wav</div> <div>▶ 0:00 / 0:04 <div></div> 🔊 ⋮</div>	3	2

All Language

▼

Evaluation

Action

Show

Delete

Show

Delete



Technologies





Items In Progress

- TTS-related reading
- Grid5000 account
 - Access to a large-scale and flexible testbed for experiment-driven research
 - Has computational resource to train the neural network models for TTS
- Evaluation approach:
 - Type: subjective
 - Participants: target min 12 people per lg
 - Languages: EN, FR
 - Modules to evaluate: speech generation, emotional contour transfer
 - Subtitles: yes for speech generation module
 - Evaluation metrics: MOS
 - Models to evaluate (candidates): 2-3
 - Time estimation: 20 min

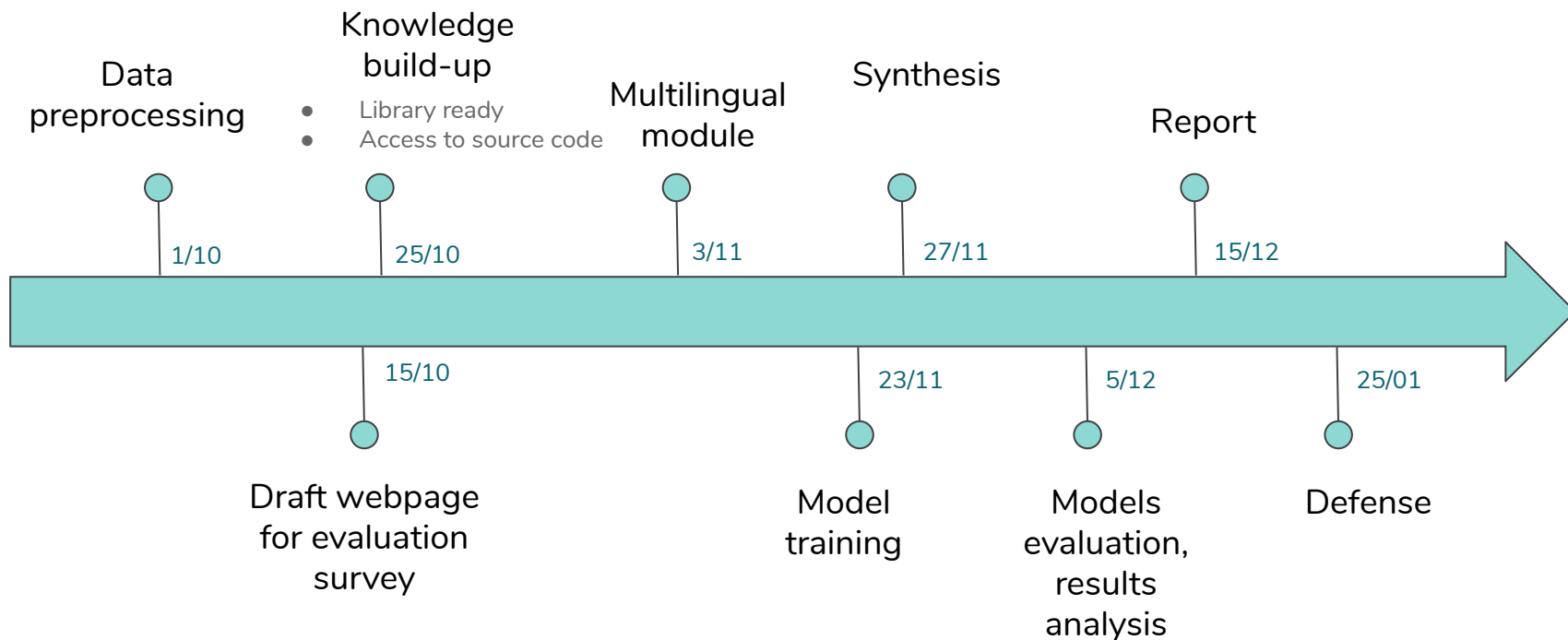


Items To Do

- Run Mellotron locally with pre-trained model
- Get acquainted with the new library and the source code
- Add multilingual module
- Train model
- Evaluate synthesized speech
- Results analysis
- Writing
 - Project report
 - Paper for Interspeech conference



Timeline



Thank you for your attention!

DO YOU HAVE ANY QUESTIONS?