

Alena YAKAVETS
Cécile MACAIRE
Chanoudom PRACH
Ludivine ROBERT

Software Project

MULTILINGUAL TEXT-TO-SPEECH SYSTEM

M2 NLP 2020-2021

December 18, 2020



Presentation Overview

- Items Done
- Items In Progress
- Challenges
- Backlog
- Timeline



Items Done

- Input file modifications.

```
/srv/storage/multispechedu@talc-data2.nancy/software_project/corpus/EmoV-DB/sam/Disgusted/Disgust_85-112_0106.wav|The emotion  
which she had suppressed burst forth now in a choking sob.|0|3|0  
/srv/storage/multispechedu@talc-data2.nancy/software_project/corpus/EmoV-DB/bea/Angry/anger_281-308_0291.wav|The weeks had gone  
by, and no overt acts had been attempted.|1|2|0  
/srv/storage/multispechedu@talc-data2.nancy/software_project/corpus/LJSpeech/wavs/LJ008-0178.wav|the bodies for identification,  
the wounded to hospitals, a cart-load of shoes, hats, petticoats, and fragments of wearing apparel were picked up.|5|0|0  
/srv/storage/multispechedu@talc-data2.nancy/software_project/corpus/SIWIS/wavs/neut_parl_s01_0346.wav|25, 9, 8, 0, 21, 7, 1, 5, 9,  
10, 14, 14, 10, 1, 9, 12, 0, 15, 15, 0, 21, 8, 14, 21, 24, 30, 0, 15, 1, 0, 5, 8, 0, 15, 0, 23, 13, 4, 3, 0, 15, 15, 27, 26, 7, 1,  
6, 4, 3, 31, 30, 24, 5, 27, 12, 24, 26, 8, 9, 21, 2, 11, 15|4|0|1  
/srv/storage/multispechedu@talc-data2.nancy/software_project/corpus/SIWIS/wavs/neut_parl_s01_0032.wav|25, 0, 15, 15, 2, 20, 22, 5,  
9, 12, 14, 26, 8, 9, 5, 14, 1, 0, 15, 10, 2, 26, 17, 8, 7, 0, 20, 7, 3, 9, 29, 31, 3, 0, 15, 1, 0, 21, 0, 8, 14, 1, 15|4|0|1  
/srv/storage/multispechedu@talc-data2.nancy/software_project/corpus/EmoV-DB/jenie/Amused/amused_29-56_0030.wav|I had faith in  
them.|3|1|0  
/srv/storage/multispechedu@talc-data2.nancy/software_project/corpus/LJSpeech/wavs/LJ023-0100.wav|that the majority were actually  
reading into the Constitution their own "personal economic predilections," and that|5|0|0  
/srv/storage/multispechedu@talc-data2.nancy/software_project/corpus/LJSpeech/wavs/LJ017-0085.wav|Palmer's plan was to administer  
poison in quantities insufficient to cause death, but enough to produce illness which would account for death.|5|0|0
```

Figure 1: Extract of the input file automatically generated by a script.

- Splitting into train (90%), test (5%) and evaluation (5%) files.
- Removing errors.



Items Done

- Move corpora (SIWIS, LJSpeech, EmoV-DB) to a shared storage space.



/srv/storage/multispechedu@talca-data2.nancy/

- Modification of the *hparams.py* file.

```
#####  
# Data Parameters      #  
#####  
load_mel_from_disk=False  
training_files='/home/cmacaire/ERISHA/filelists_2/English_French_train.txt'  
validation_files='/home/cmacaire/ERISHA/filelists_2/English_French_eval.txt'  
text_cleaners=['english_cleaners']
```

```
#####  
# Model Parameters    #  
#####  
n_symbols=len(symbols) + 37
```

```
#####  
# Reference Encoder Network Hyperparameters #  
#####  
  
speaker_encoder_type = 'vae'  
expressive_encoder_type = 'vae'  
language_encoder_type = 'vae'  
  
emotion_classes = 5  
speaker_classes = 6  
language_classes = 2
```

Figure 2: Modifications made on parameters from *hparams.py* file.

Items In Progress - Training

- Models:
 - GST - *Global Style Tokens*
 - VAE - *Variational Autoencoder*
 - GMVAE - *Gaussian Mixture VAE*
 - X-vector
- Book GPU nodes on Grid5000:
 - training time set to 24 hours.
 - passive mode.

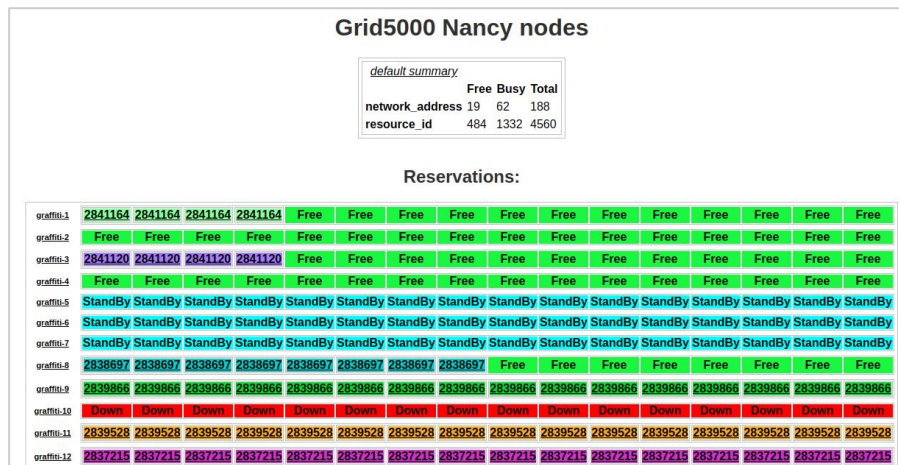


Figure 3: Grid5000 Nancy nodes webpage.

2841259	cprach	Running production	-I "(((type = 'default') AND (type != 'default' OR max_walltime >= 86400 OR max_walltime <= 0)) AND production = 'YES') AND exotic = 'NO'}host=1,walltime=24:0:0"	PASSIVE	(cluster='grele' or cluster='graffiti' or cluster='grue') AND maintenance = 'NO'	None	24:0:0	2020-12-17 14:23:14	2020-12-17 14:23:15	2020-12-17 14:23:15
2841309	lrobert	Running production	-I "(((type = 'default') AND (type != 'default' OR max_walltime >= 86400 OR max_walltime <= 0)) AND production = 'YES') AND exotic = 'NO'}host=1,walltime=24:0:0"	PASSIVE	(cluster='grele' or cluster='graffiti' or cluster='grue') AND maintenance = 'NO'	None	24:0:0	2020-12-17 14:43:36	2020-12-17 14:43:49	2020-12-17 14:43:49

Figure 4: Example of two training processes on Grid5000 nodes.



Items In Progress - Training

- Model saved in a checkpoint file each n steps.
- Stdout file per jobs (epoch, train loss, validation loss information).

```
iteration 110000
Epoch: 4
Train loss 110001 0.120133 Grad Norm 1.175963 1.35s/it
Train loss 110002 1.169019 Grad Norm 3.049260 1.80s/it
Train loss 110003 1.202280 Grad Norm 5.466412 0.65s/it
Train loss 110004 0.097937 Grad Norm 0.652959 1.67s/it
Train loss 110005 0.086411 Grad Norm 0.799507 2.83s/it
Train loss 110006 0.079062 Grad Norm 0.596764 1.46s/it
Train loss 110007 1.312290 Grad Norm 3.029655 2.50s/it
Train loss 110008 1.065159 Grad Norm 1.890132 0.84s/it
Train loss 110009 1.300889 Grad Norm 3.178862 0.97s/it
Train loss 110010 1.185444 Grad Norm 1.625215 2.30s/it
Train loss 110011 1.139958 Grad Norm 1.680999 1.44s/it
Train loss 110012 0.143788 Grad Norm 1.918057 2.64s/it
Train loss 110013 1.177848 Grad Norm 2.579276 1.61s/it
Train loss 110014 0.132099 Grad Norm 1.977031 2.15s/it
Train loss 110015 1.175572 Grad Norm 1.430215 2.90s/it
Train loss 110016 1.177395 Grad Norm 2.606436 1.07s/it
Train loss 110017 1.220565 Grad Norm 1.918178 1.47s/it
Train loss 110018 1.242177 Grad Norm 1.628366 3.05s/it
Train loss 110019 1.121986 Grad Norm 3.119315 0.76s/it
Train loss 110020 0.103050 Grad Norm 0.986830 0.90s/it
Train loss 110021 0.122215 Grad Norm 0.915862 1.10s/it
Train loss 110022 1.172873 Grad Norm 2.747619 0.94s/it
```

```
Validation loss 120000: 4.592939
Saving model and optimizer state at iteration 120000 to /srv/
storage/multispechedu@talca-data2.nancy/software_project/vae/
output/checkpoint_120000
Train loss 120001 0.182466 Grad Norm 2.207243 1.26s/it
```

Figure 3: Stdout file generated during the training of a model.



Challenges

About the data...

Some issues (corpus, processed data)



duplicates,



no transcript,



etc.

About the training...

Model

- Sometimes a model would have trained for 8 hours before an error occurs and the process gets terminated.

Time

- The models require extensive train time (2 weeks each).

Grid5000

- No notification if the training stops due to an error.
- Due to the access level policy in Grid5000 we cannot currently run the training in autonomous mode for more than 24h (expected max - 168h).
- Issues preventing from training the model in autonomous mode (now resolved).

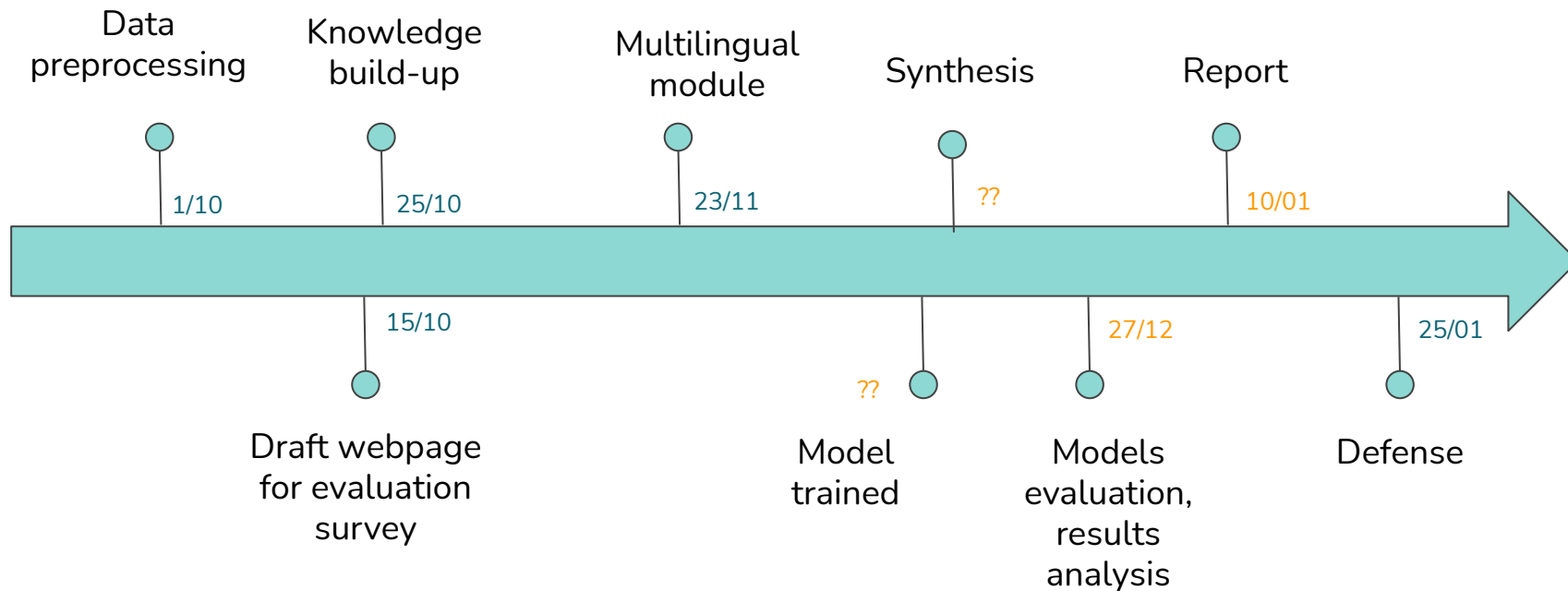


Items To Do

1. Evaluate synthesized speech:
 - Synthesize speech samples for each trained model,
 - Upload samples to the website,
 - Call-to-action for participants.
2. Results analysis & interpretation.
3. Writing:
 - Paper for Interspeech conference,
 - Final report.



Timeline



Thank you for your attention!

DO YOU HAVE ANY QUESTIONS?