TTS

Albert, Wenjun, Soklay, Shalini

Server
:5000
Gunicorn
:5000
:5000
Request
Response
Request
Response
UI
:3000
Response
Request
:80
Proxy
:80
Response
Request

Flexibility

- add load balancing to proxy
- if there's need for more containers (high traffic)
  - spin up more server containers: **--scale tts-api=4** or automatic orchestration
- swapping server without service down-time

Done / progress

- Language classification model
- English, French TTS models
- Server, ui with integration through proxy in the same hidden network
  - server exposes port 5000 through gunicorn
  - nginx proxy works on port 80 (default HTTP port)
  - ui exposes port 3000 and communicates with the server through the proxy
- Server generates a wav file that is being sent to ui running in user's browser
  - play the sound
- Initial analyses
- Model evaluation
  - Amazon Turk crowdsourcing
  - Online questionnaire
- System deployment
  - problem - one rather big model added through git lfs; is it going to work?

French Synthesis Demo

Human Evaluation

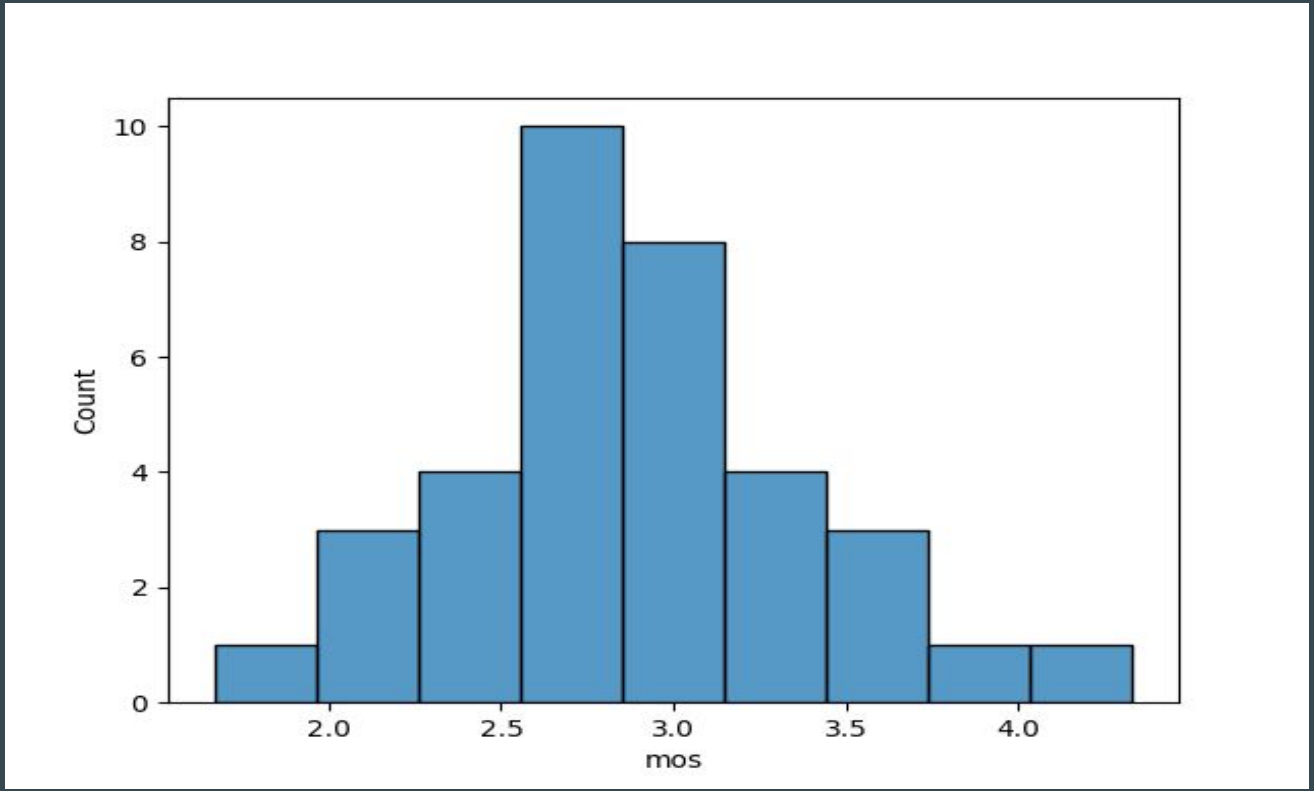
Dataset: Voice Conversion Challenge 2018

35 audios

3 evaluators for 1 audio

use average as final grade

Human Evaluation



Auto Evalution

AutoMos & MosNet

Dataset: VCC 2018

20k+ (audio & mos_grade)

input: spetrogram

structure: CNN + LSTM

Evaluation: Intelligibility

1. Objectivity

+ Google ASR
+ WER

```
hypo = []
for f in file:
    transcription = transcribe_audio(f)
    hypo.append(transcription)
```

```
hypo
```

```
['National Gallery gets all clear',
 'community to help homeless',
 'death toll continues to climb in South Korean sub',
 'what time is allocated for domestic violence victims',
 'Korean subway fire a314 still miss']
```

```
ground_truth = ["National Gallery gets all clear", "community to help homeless",
```

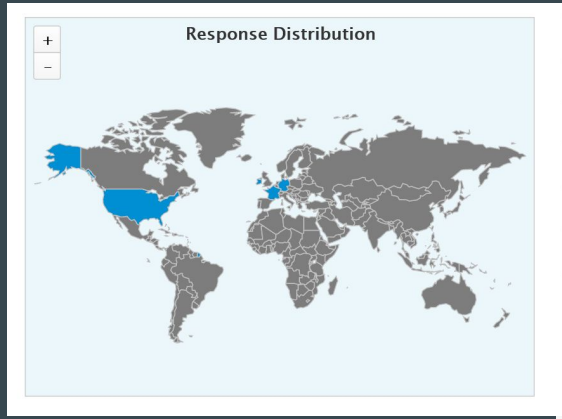
```
#The most simple use-case is computing the edit distance between two strings:
from jiwer import wer
error = wer(ground_truth, hypo)
```

```
error
```

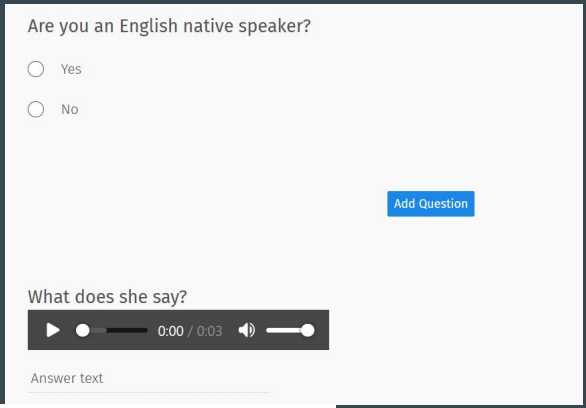
```
0.03125
```

Evaluation: Intelligibility

1. Subjectivity (10- 15 English Native Speakers)



Countries	Responses
FR	66.67%
US	22.22%
IE	7.41%
DE	3.70%
Total	100.00%



01/30/2022	65330228	National gallery gets all clear after
01/30/2022	65330103	National gallery gets all clear after
01/29/2022	65280832	National gallery gets all clear after
01/29/2022	65276508	National gallery gets all clear after
01/29/2022	65273357	National Gallery gets all clear after
01/29/2022	65271883	National gallery gets all clear after
01/29/2022	65267900	National gallery gets all clear after

Next Steps

Fine tune the model

Evaluation (French)

Beautify the Web Application