

TEXT TO SPEECH

AUDIO

Jorge Muñoz Aguado

INDEX

1. Introduction to TTS

2. ESPnet

3. Results

4. Conclusion and resources

INTRODUCTION TO TTS

Hello, how are you?

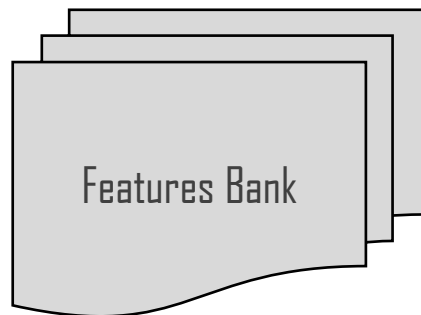


INTRODUCTION TO TTS (I)

Hello, how are you?

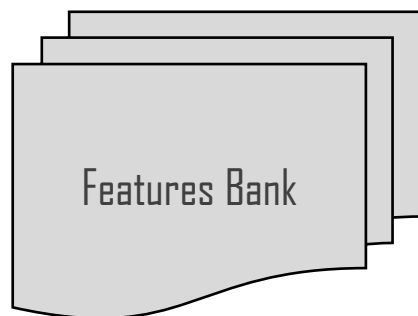


Feature extraction

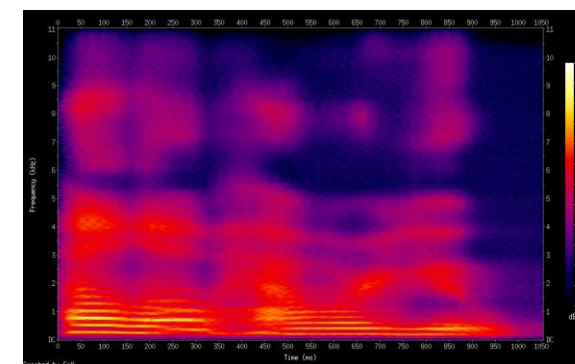


INTRODUCTION TO TTS (2)

Hello, how are you?



Generator model

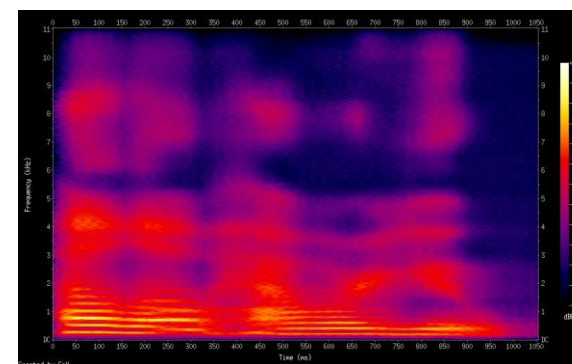
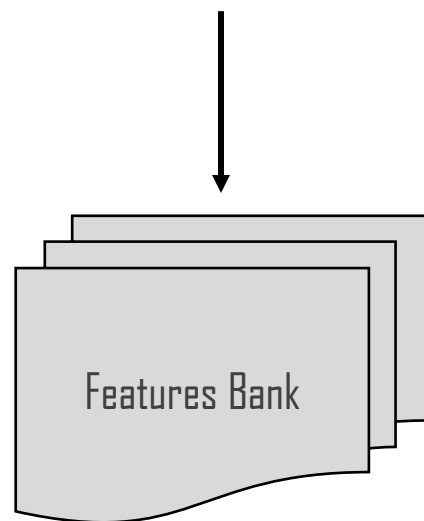


INTRODUCTION TO TTS (3)

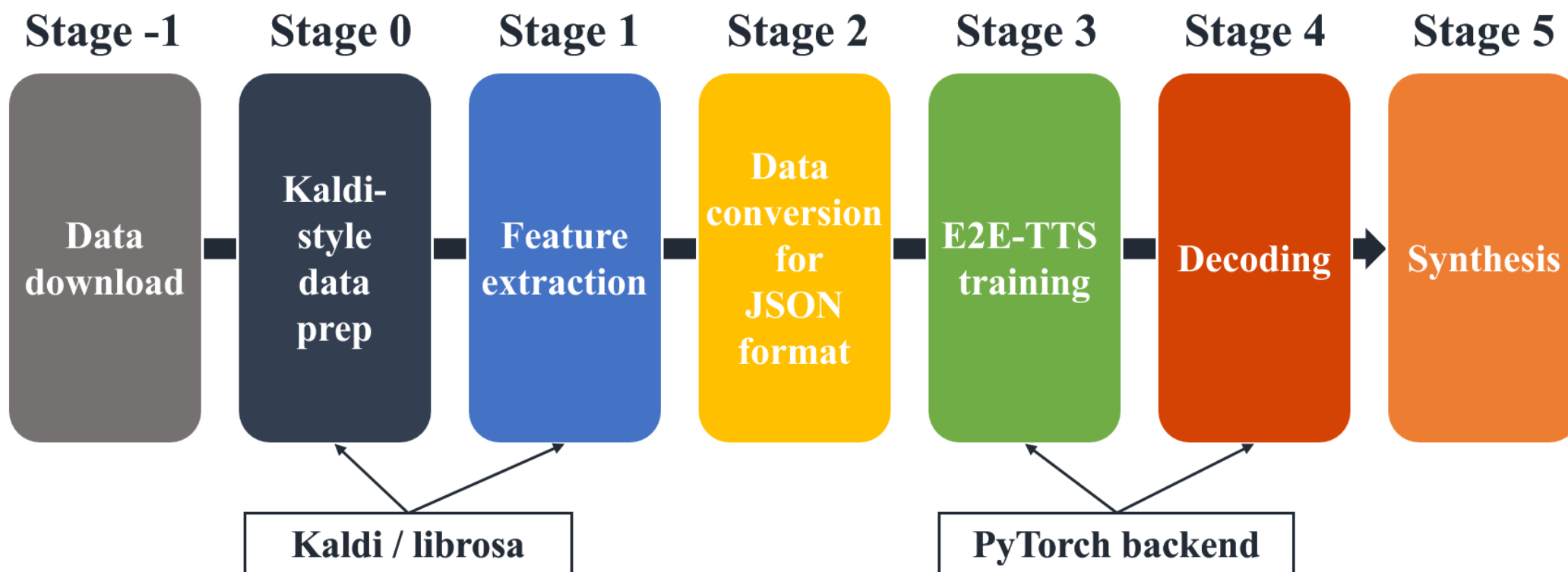
Hello, how are you?



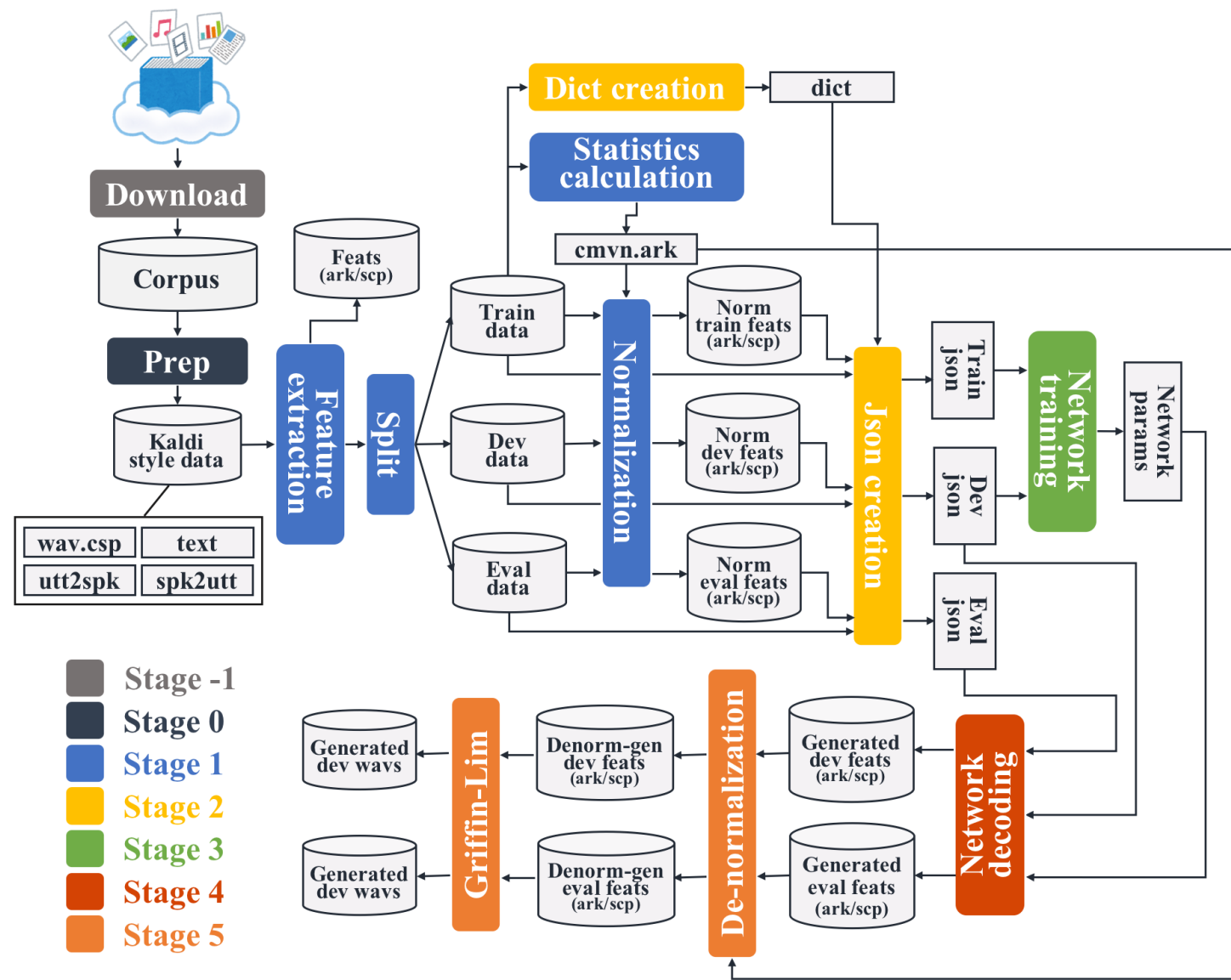
Vocoder



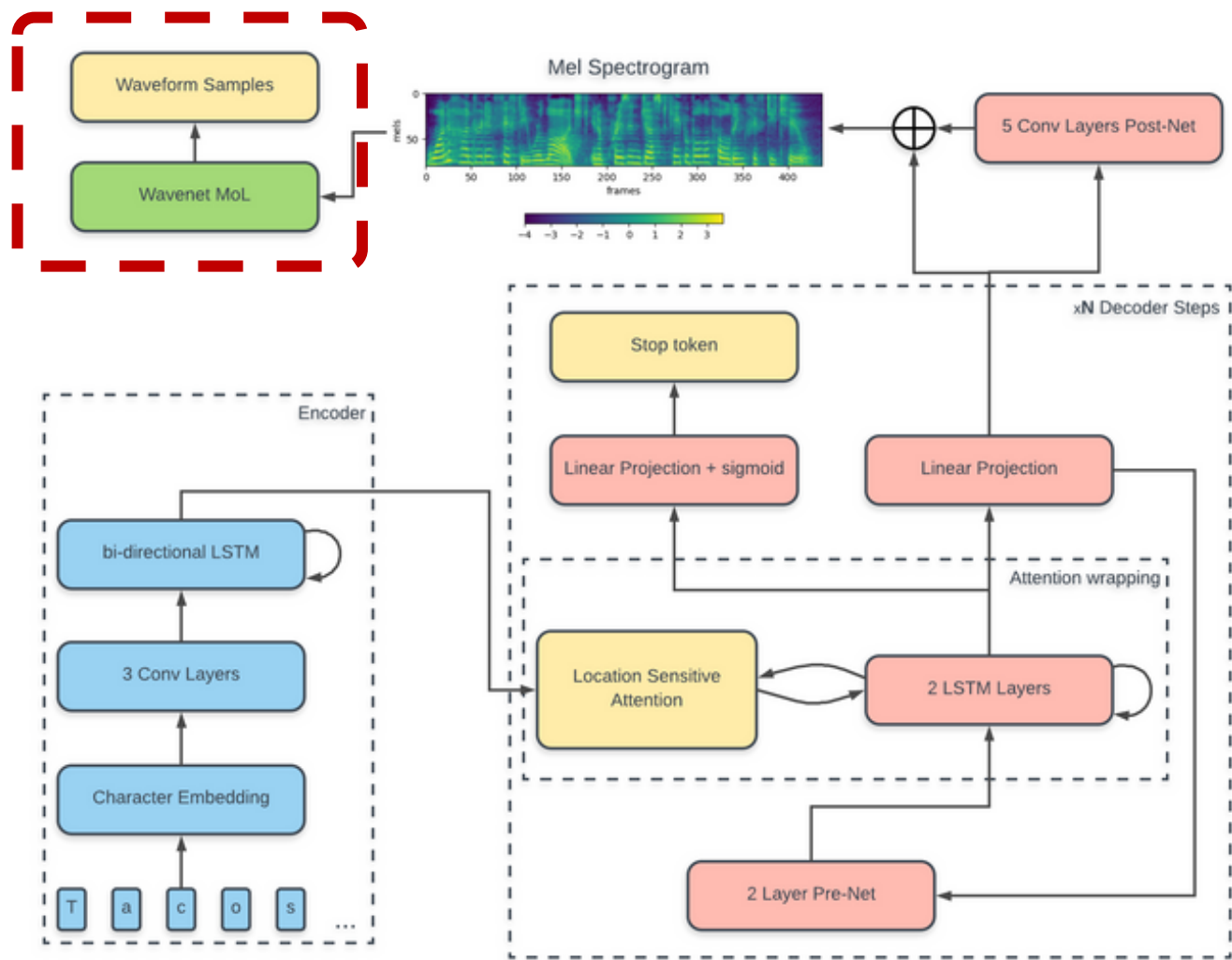
ESPNET



ESPNET



ESPNET (TACOTRON.2)

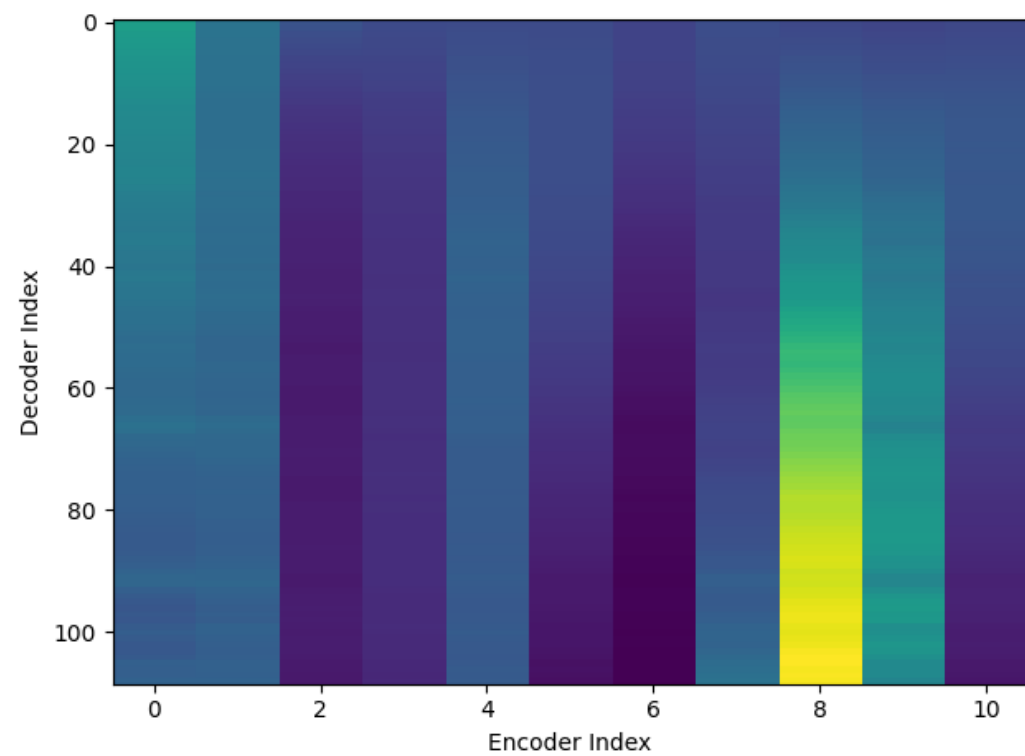
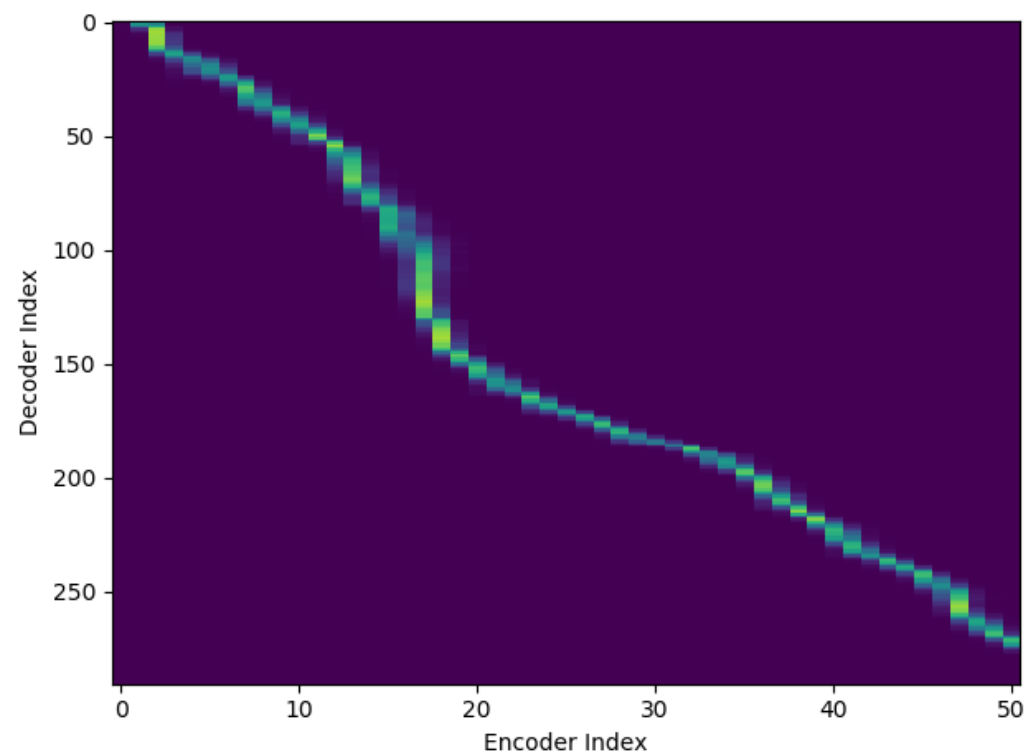


RESULTS (EVALUATION)

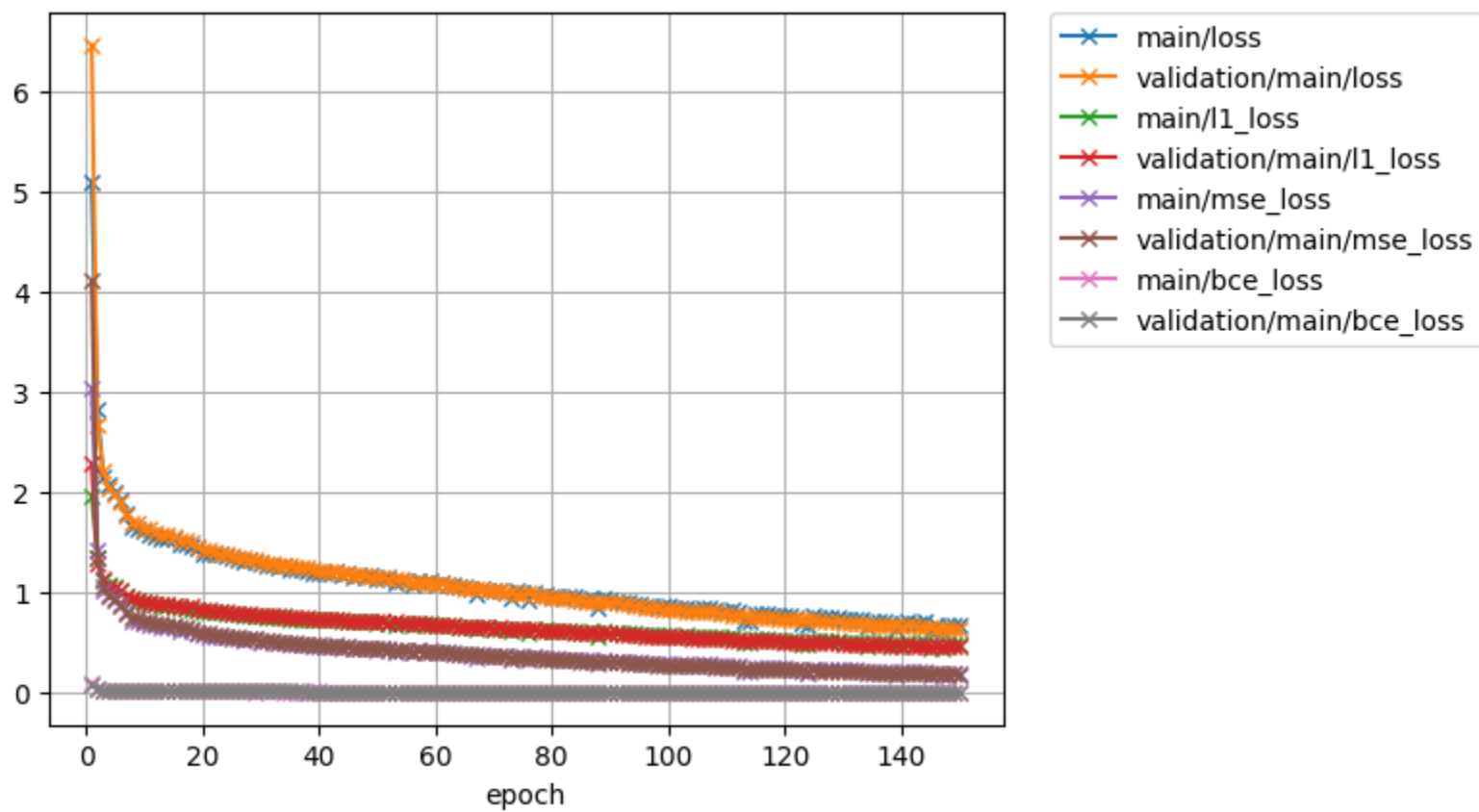
Subjective evaluation: [MOS](#) (Mean Opinion Score)

Objective evaluation: [MCD](#) (Mel Cepstral Distortion)

RESULTS



RESULTS



RESULTS (Trained model)

tacotron2

- *Based on Attention Network*



JANUARY SEVENTH NINETEEN SIXTY SEVEN



Yes



No



M A R G A R E T M O R R I S O N

RESULTS (Pre-trained model)

Hello, how are you?



`ljspeech.fastspeech.v1`



`ljspeech.tacotron2.v3`



`ljspeech.tacotron2.v3 + Vocoder`

CONCLUSIONS

```
pip install --upgrade google-cloud-texttospeech
```

```
pip install --upgrade google-cloud-speech
```

RESOURCES

https://github.com/JorgeMunnozAguado/Audio_Project

<https://github.com/espnet/interspeech2019-tutorial>

(TTS) <https://cloud.google.com/text-to-speech/docs/libraries#client-libraries-install-python>

(ASR) <https://cloud.google.com/speech-to-text/docs/libraries?hl=es-419>

Shen, Jonathan, et al. "Natural tts synthesis by conditioning wavenet on mel spectrogram predictions." *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018.