TEXT TO SPEECH

AUDIO

Jorge Muñoz Aguado

INDEX

1. Introduction to TTS

2. ESPnet

3. Results

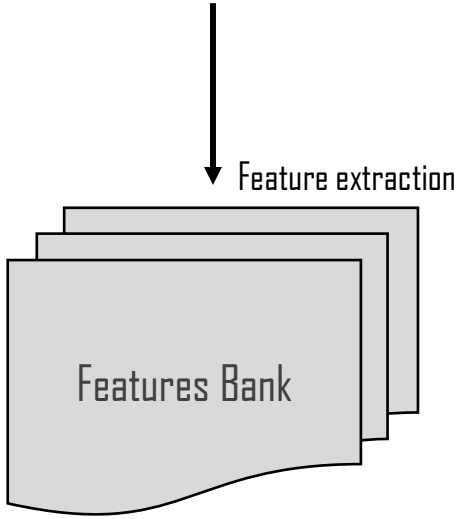
4. Conclusion and resources

INTRODUCTION TO TTS

Hello, how are you?

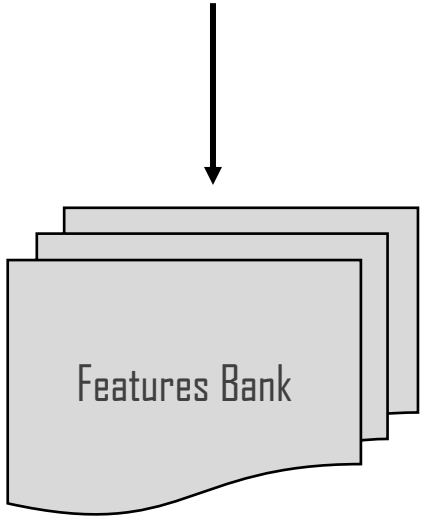
INTRODUCTION TO TTS (1)

Hello, how are you?

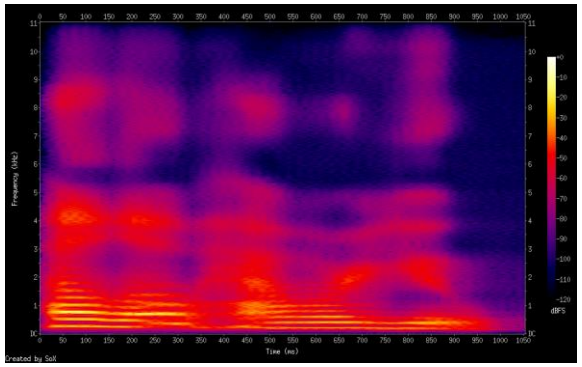


INTRODUCTION TO TTS (2)

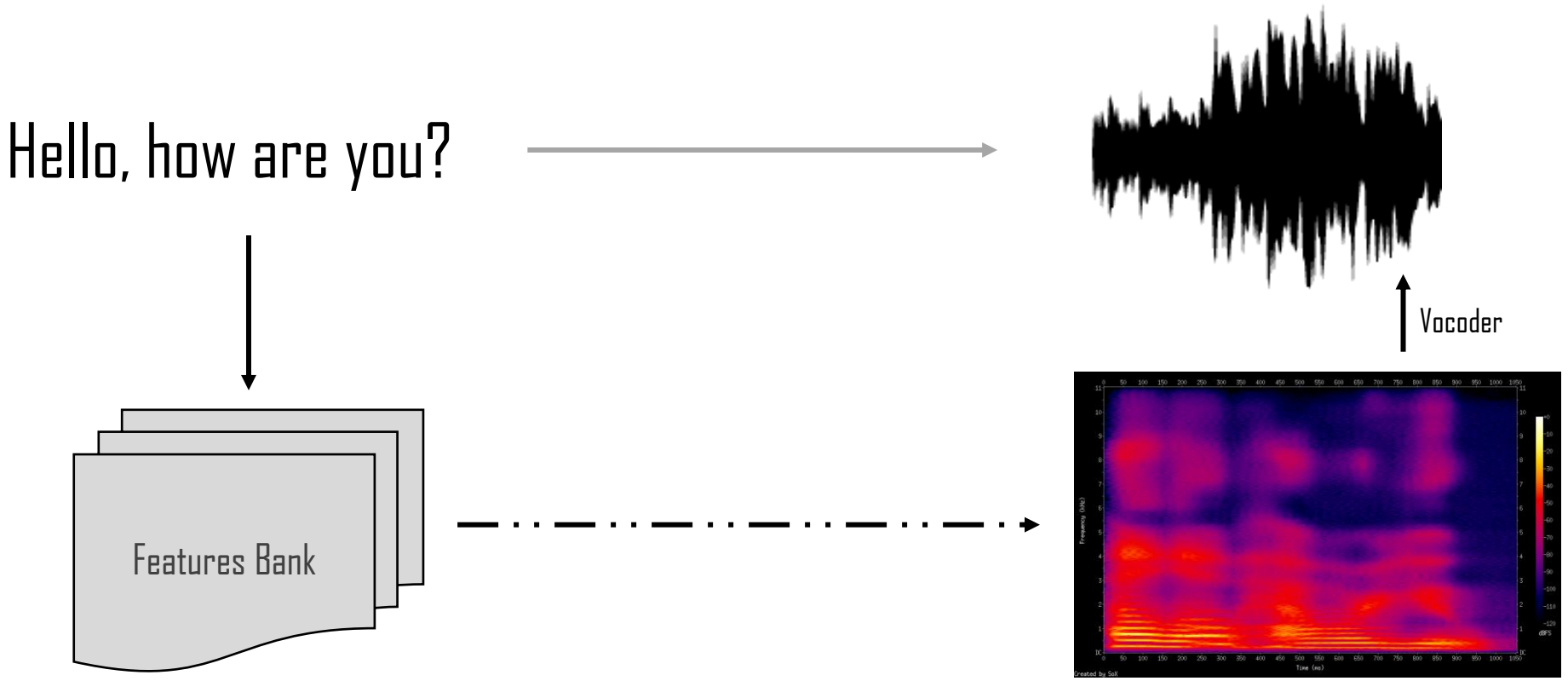
Hello, how are you?



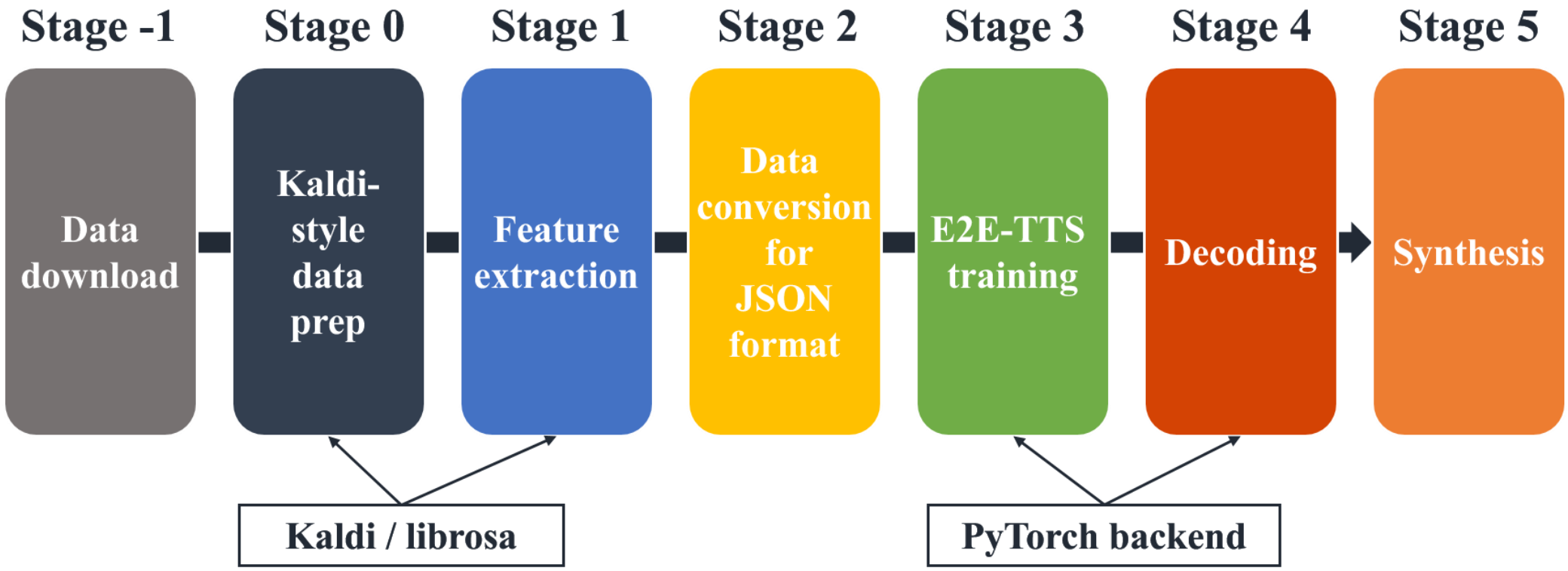
Generator model



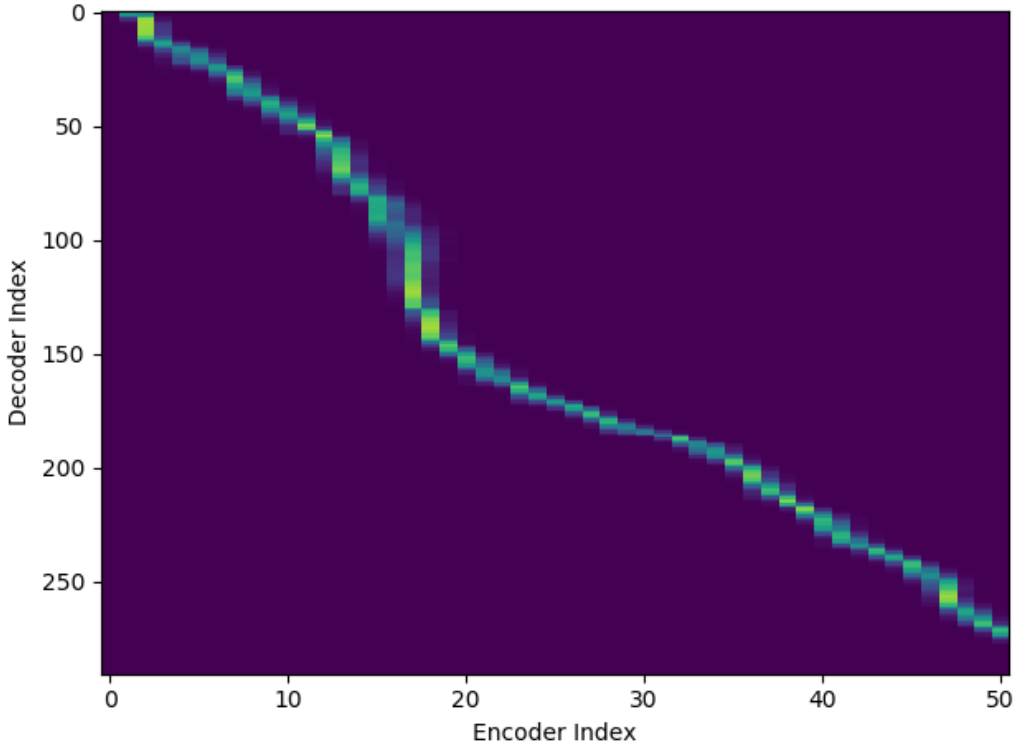
INTRODUCTION TO TTS (3)

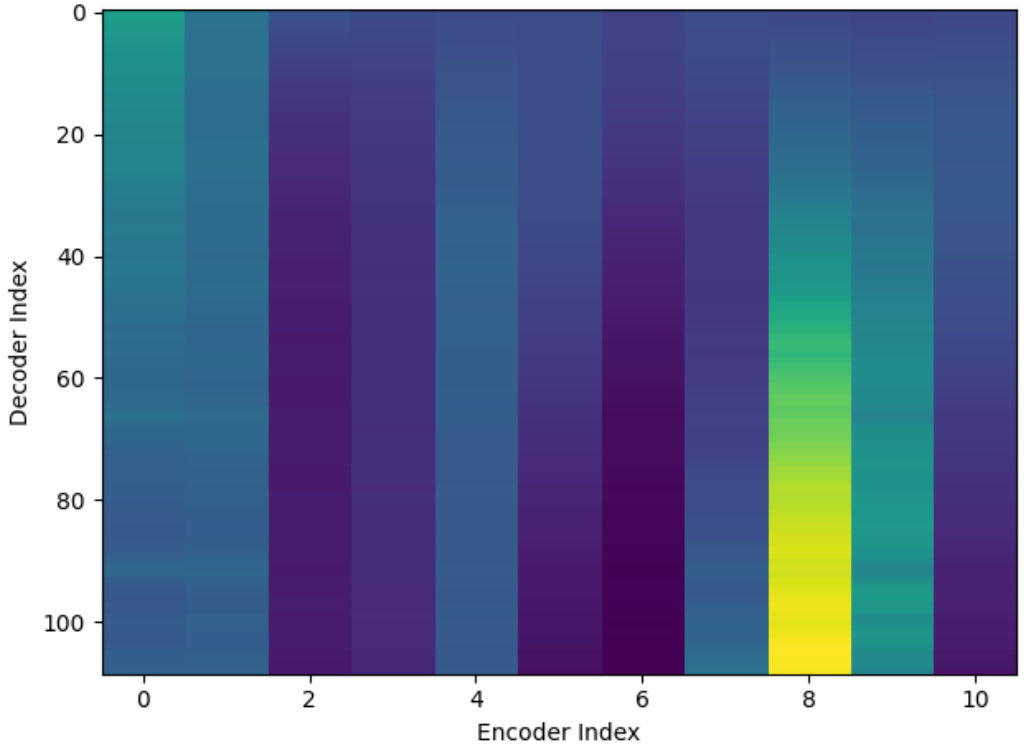


ESPNET

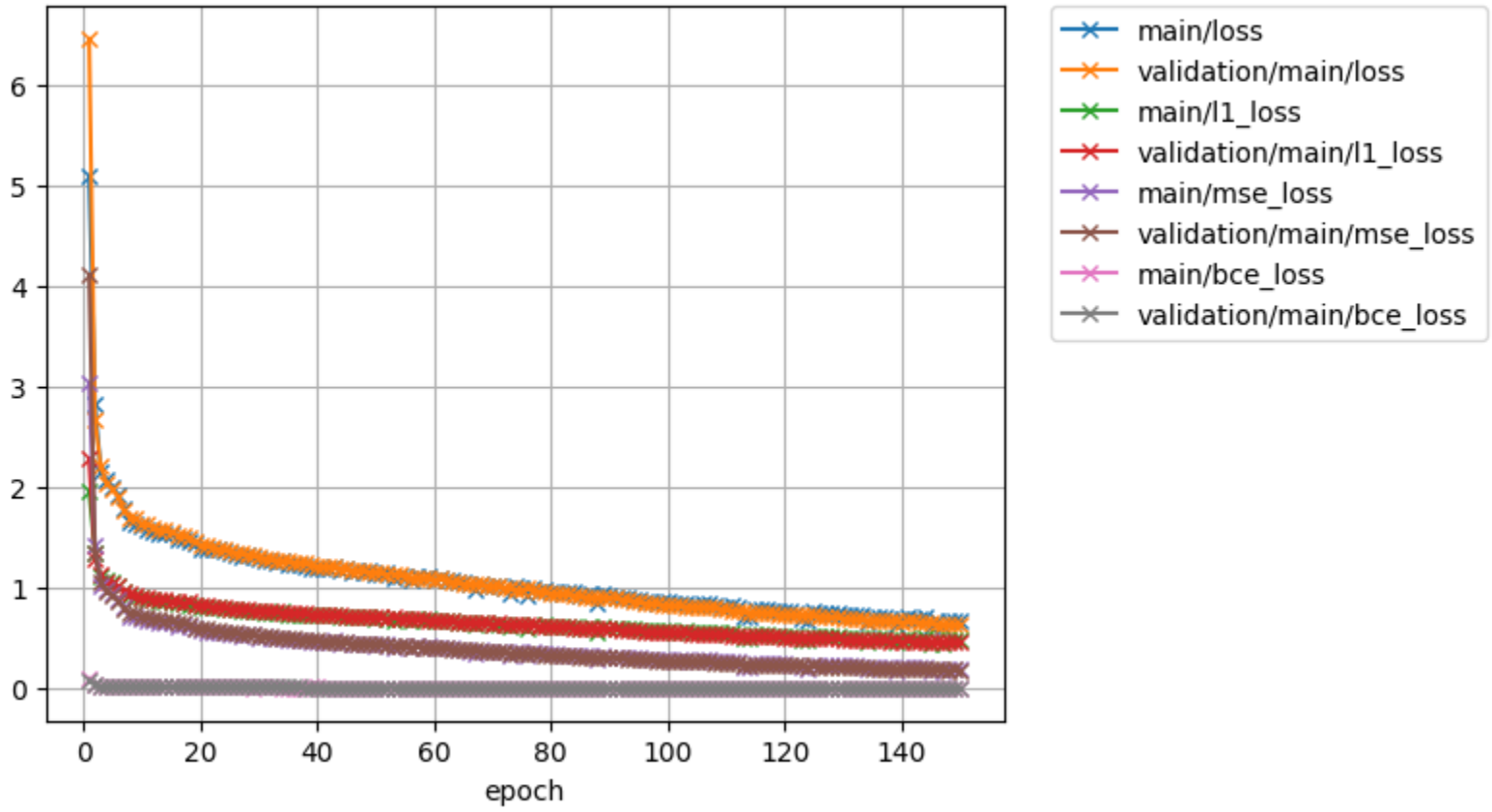

Stage -1
Stage 0
Stage 1
Stage 2
Stage 3
Stage 4
Stage 5
Data download
Kaldi-style data prep
Feature extraction
Data conversion for JSON format
E2E-TTS training
Decoding
Synthesis
Kaldi / librosa
PyTorch backend

RESULTS


Decoder Index
Encoder Index
0
50
100
150
200
250
0
10
20
30
40
50


Decoder Index
Encoder Index
0
20
40
60
80
100
0
2
4
6
8
10

RESULTS


main/loss
validation/main/loss
main/l1_loss
validation/main/l1_loss
main/mse_loss
validation/main/mse_loss
main/bce_loss
validation/main/bce_loss
0
1
2
3
4
5
6
0
20
40
60
80
100
120
140
epoch

RESULTS (Trained model)

tacotron2.v4 - *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 colab

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