

# Guarani

Using pytorch deepspeech

- [Data stats](#)
- [Conversion of sph files](#)
- [New Page](#)

# Data stats

Conversational training: 37.55 hours

```
find wav/ | xargs -I {} -r soxi -D {} | awk '{sum += $1} END {print sum}'
```

Scripted training: 12.82 hours (using same command in the scripted training dir)

# Conversion of sph files

```
(metalearn) (base) kenneth@mwanafunzi:/home/data/corpora/speech/IARPA-babel305b-v1.0c-build$ find -
```

# New Page

Early stuff I was doing was seeing the first epochs perform the best and high sensitivity to number of rnn layers (less was performing better) this ended up being because the gradients were exploding.

To solve this, I specified `--max-norm 1` which forces gradient clipping.

Ended up getting 66 CER after 10 epochs.

Trying again with 100 epochs and data augmentation (pitch perturbation and gain modulation).

```
python train.py --train-manifest data/train_combined.csv --val-manifest
data/_home_data_corpora_speech_IARPA-babel305b-v1.0c-
build_converted_BABEL_OP3_305_conversational_dev__manifest.csv --model-path
models/100epoch2.pth --max-norm 1 --epochs 100 --opt-level 00 --cuda --augment --labels-path
data/BABEL/guarani_labels.json --tensorboard
```

```
python train.py --train-manifest data/train_combined.csv --val-manifest
data/_home_data_corpora_speech_IARPA-babel305b-v1.0c-
build_converted_BABEL_OP3_305_conversational_dev__manifest.csv --model-path
models/100epoch2.pth --max-norm 50 --epochs 100 --opt-level 00 --cuda --augment --labels-path
data/BABEL/guarani_labels.json --tensorboard --augment
```

# 1 gru successful

```
python train.py --train-manifest data/train_combined.csv --val-manifest
data/_home_data_corpora_speech_IARPA-babel305b-v1.0c-
build_converted_BABEL_OP3_305_conversational_dev__manifest.csv --model-path
models/1_hi_aug.pth --max-norm 100 --hidden-layers 1 --epochs 20 --opt-level 00 --cuda --
labels-path data/BABEL/guarani_labels.json --loss-scale 1 --tensorboard --id
1_hid_max_norm_100
```

result: WER 88.366 CER 43.116

still achieving better performance  
after 20 epochs, so I should rerun  
with more epochs

## 2 gru

```
python train.py --train-manifest data/train_combined.csv --val-manifest
data/_home_data_corpora_speech_IARPA-babel305b-v1.0c-
build_converted_BABEL_OP3_305_conversational_dev__manifest.csv --model-path
models/2_hi_aug.pth --max-norm 100 --hidden-layers 2 --epochs 40 --opt-level 00 --cuda --
labels-path data/BABEL/guarani_labels.json --loss-scale 1 --tensorboard --id
2_hid_max_norm_100
```

## results with max norm = 100

Average WER = 74.590 Average CER = 29.756

Going to try again with higher max norm (default 400) and see how that goes. Before the gradients were exploding when I did that.

## 3 gru

max norm = 100, after 40 epochs: WER = 60.519 CER = 21.876

## 4 gru

max norm = 100, after 40 epochs: WER = 52.413 CER = 17.833

## 5 gru

(ran on qivalluk). 40 epochs, Average WER 14.239 Average CER 3.655. Used nearly all default arguments.