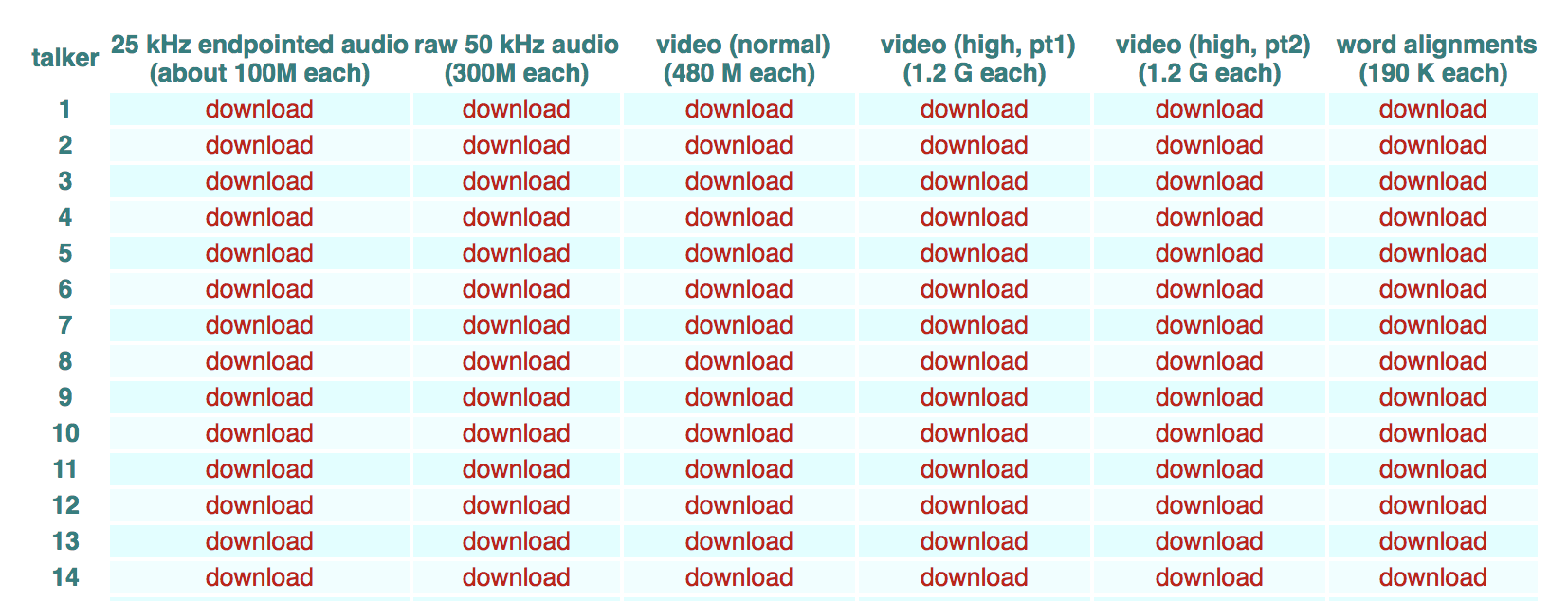
Digit number recognition on Grid audiovisual sentence corpus dataset

Main Steps:

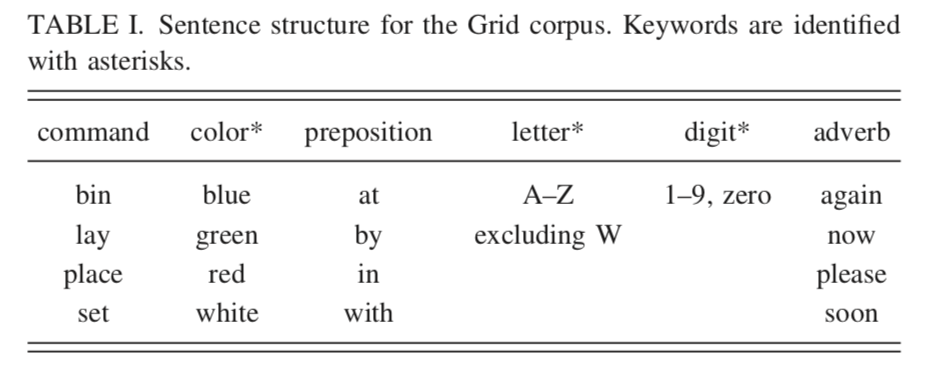
1. Segment the audio files
2. Generate spectrogram
3. Classify the spectrogram using CNN

Step1: Audio files Segmentation

Segment the digit number audio recordings from the raw 50 KHz audio files according to the word alignments files

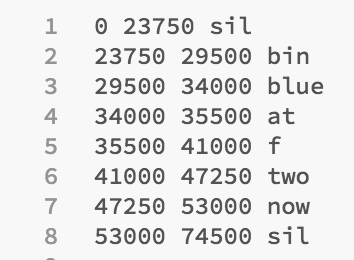


The number of talker in the dataset is 34 and the number of sentences per talker is 1,000. The sentence consists of six components, such as ‘bin blue at f two now’. The format of the sentence is shown as follows:



The audio recordings can be segmented according to the word alignments files.

The format of the word alignments files is shown as follows:

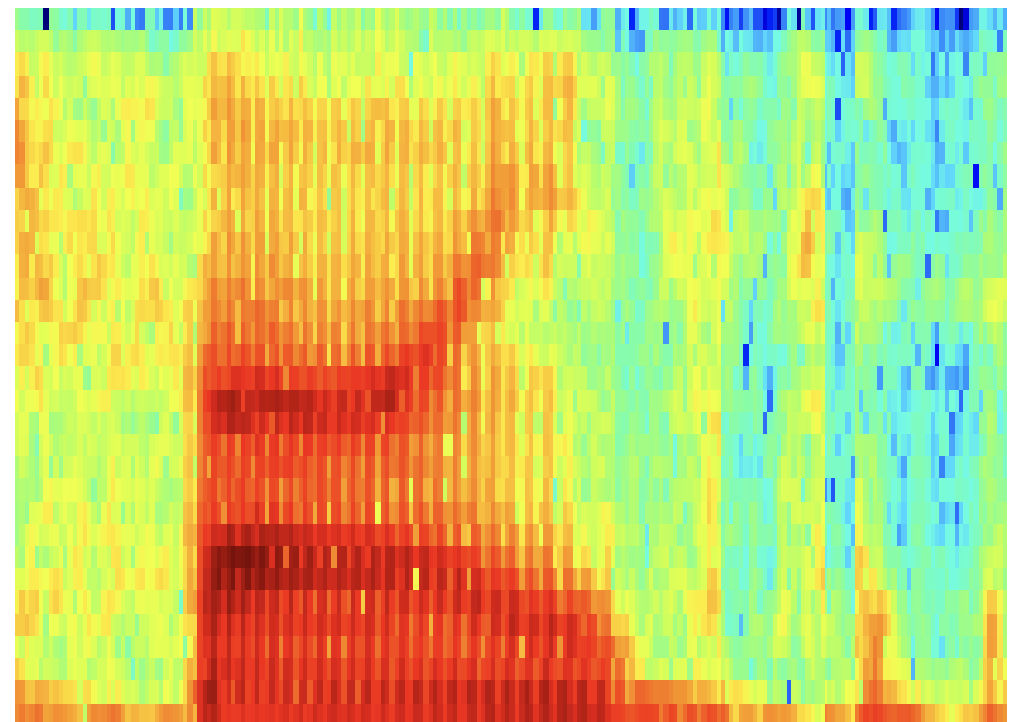
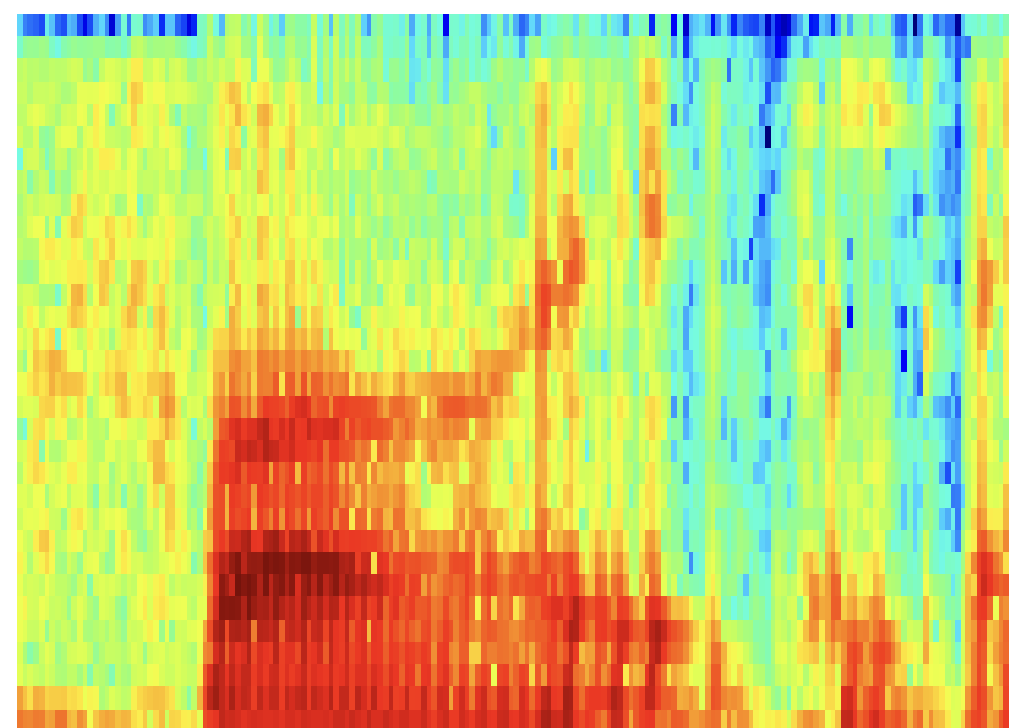
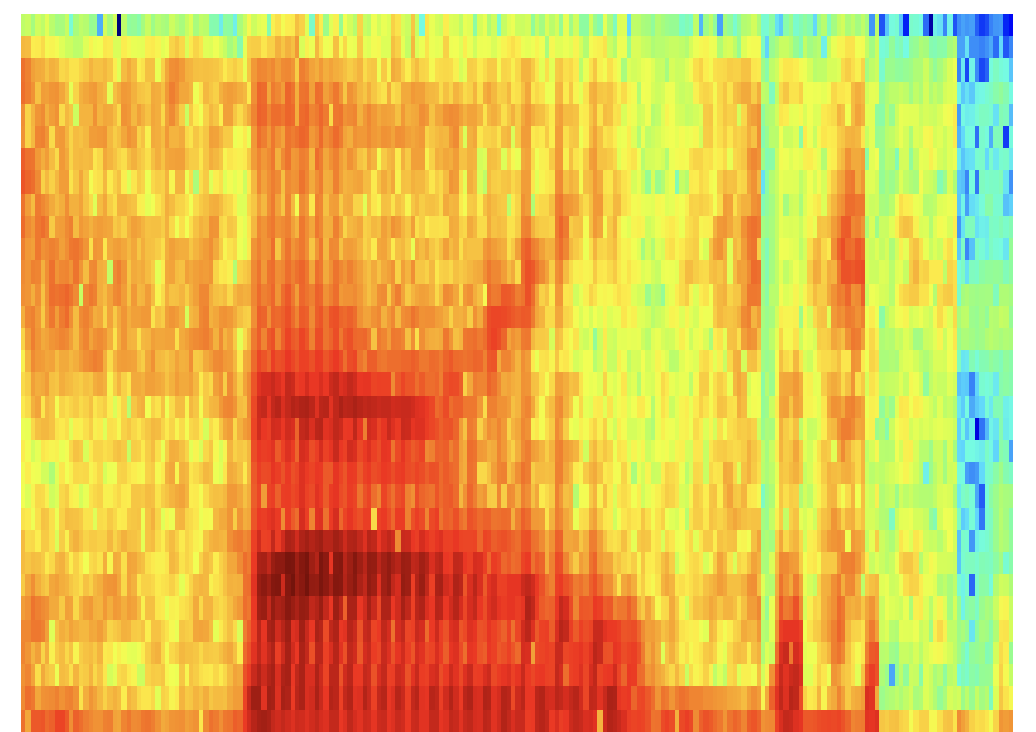


Step2: Generate spectrogram

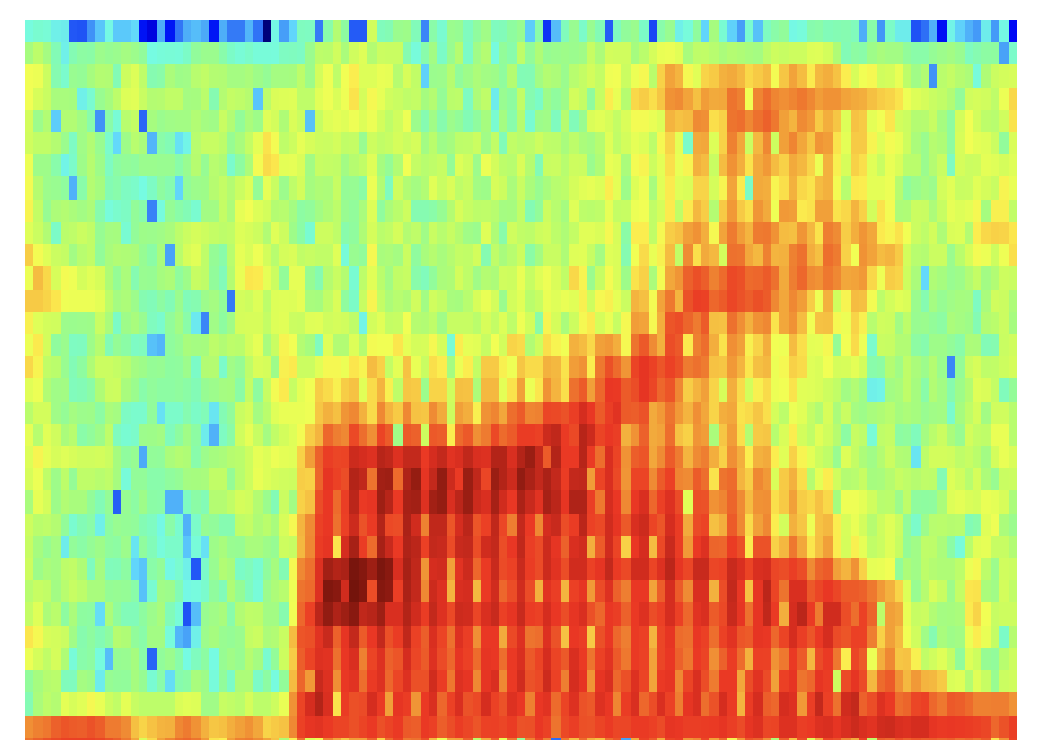
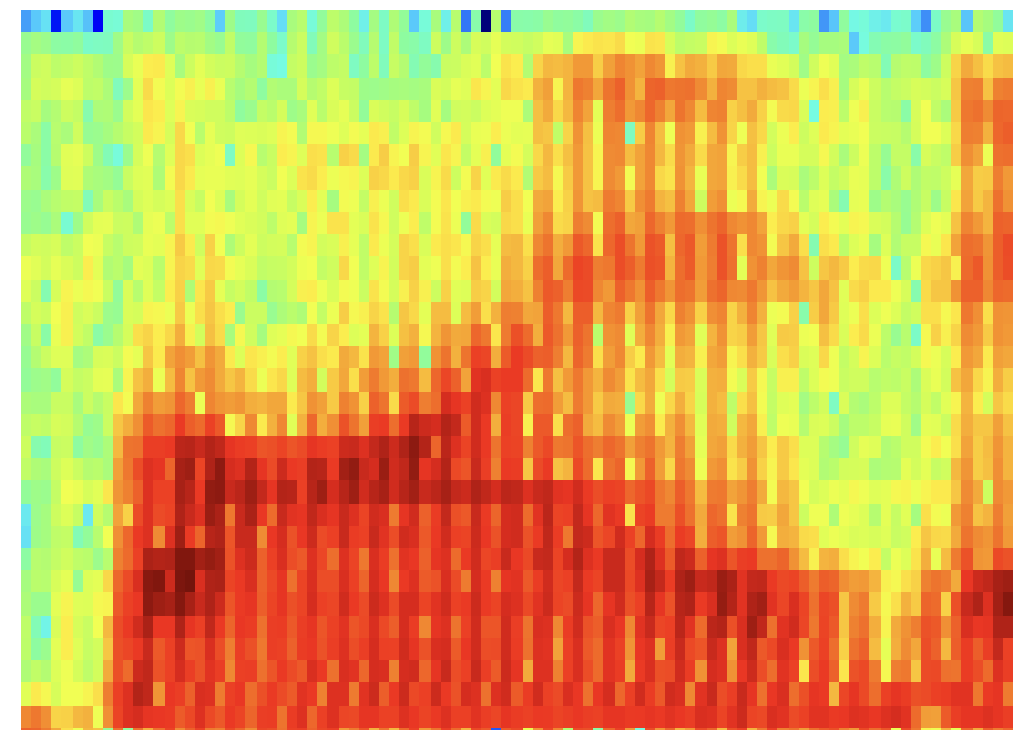
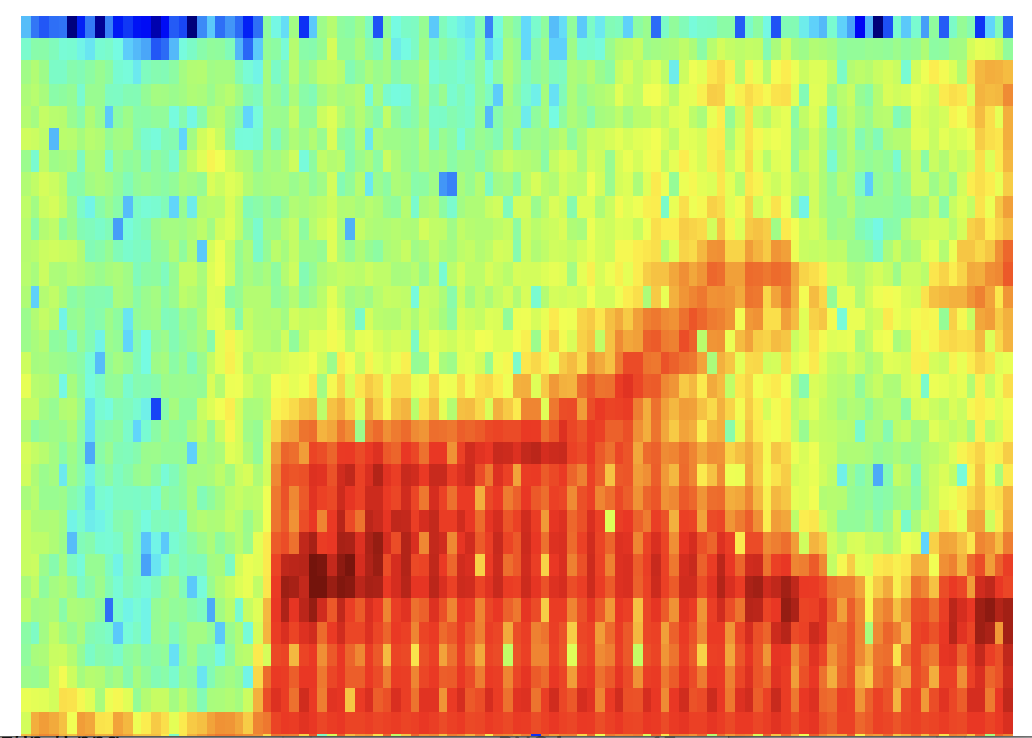
The audio sample is converted to spectrogram with Hamming window of 64 samples duration and 3/4 overlap.

Some spectrogram examples (digit 5):

Speaker3:



Speaker5:



Time performance of Step 1 and Step 2:

About 350s for each speaker (1,000 audio recordings)

Step3: Classify the spectrum using CNN (implemented in TensorFlow)

Soft

max

full

2

full

1

pool

2

conv 2

conv

1

pool 1

Filter in Convolutional Layer: [3, 3, 3, 16]

Strides in Convolutional Layer: [1, 1, 1, 1]

Strides in Max Pooling Layer: [1, 2, 2, 1]

Number of neural in Fully Connected Layer: 128

Learning rate: 0.005

Number of Steps: 2000

Batch Size: 128

Device: Pelican Server (with GPU)

Dataset: Speaker 3 and 5 (80% for training and 20% for testing randomly)

Time Performance: about 30 mins

Results:

Step 0,train loss = 2.3066,train accuracy = 12.50%,test accuracy = 10.16%

Step 50,train loss = 0.5256,train accuracy = 85.94%,test accuracy = 73.44%

Step 100,train loss = 0.1428,train accuracy = 96.88%,test accuracy = 85.16%

Step 150,train loss = 0.1567,train accuracy = 94.53%,test accuracy = 81.25%

Step 200,train loss = 0.0523,train accuracy = 98.44%,test accuracy = 85.94%

Step 250,train loss = 0.0276,train accuracy = 99.22%,test accuracy = 85.94%

Step 300,train loss = 0.0028,train accuracy = 100.00%,test accuracy = 89.06%

Step 350,train loss = 0.0294,train accuracy = 98.44%,test accuracy = 88.28%

Step 400,train loss = 0.0237,train accuracy = 99.22%,test accuracy = 82.81%

Step 450,train loss = 0.0025,train accuracy = 100.00%,test accuracy = 87.50%

Step 500,train loss = 0.0154,train accuracy = 99.22%,test accuracy = 85.16%

Step 550,train loss = 0.0336,train accuracy = 98.44%,test accuracy = 85.94%

Step 600,train loss = 0.0028,train accuracy = 100.00%,test accuracy = 92.19%

Step 650,train loss = 0.0285,train accuracy = 99.22%,test accuracy = 82.03%

Step 700,train loss = 0.0085,train accuracy = 100.00%,test accuracy = 86.72%

Step 750,train loss = 0.0275,train accuracy = 98.44%,test accuracy = 92.97%

Step 800,train loss = 0.0036,train accuracy = 100.00%,test accuracy = 89.84%

Step 850,train loss = 0.0235,train accuracy = 99.22%,test accuracy = 89.84%

Step 900,train loss = 0.0185,train accuracy = 99.22%,test accuracy = 88.28%

Step 950,train loss = 0.0378,train accuracy = 98.44%,test accuracy = 91.41%

Step 1000,train loss = 0.0083,train accuracy = 100.00%,test accuracy = 85.16%

Step 1050,train loss = 0.0234,train accuracy = 99.22%,test accuracy = 89.06%

Step 1100,train loss = 0.0086,train accuracy = 100.00%,test accuracy = 91.41%

Step 1150,train loss = 0.0154,train accuracy = 99.22%,test accuracy = 85.94%

Step 1200,train loss = 0.0328,train accuracy = 96.88%,test accuracy = 85.94%

Step 1250,train loss = 0.0026,train accuracy = 100.00%,test accuracy = 89.06%

Step 1300,train loss = 0.0238,train accuracy = 99.22%,test accuracy = 87.50%

Step 1350,train loss = 0.0295,train accuracy = 98.44%,test accuracy = 87.50%

Step 1400,train loss = 0.0218,train accuracy = 99.22%,test accuracy = 89.84%

Step 1450,train loss = 0.0745,train accuracy = 97.66%,test accuracy = 87.50%

Step 1500,train loss = 0.2162,train accuracy = 93.75%,test accuracy = 85.16%

Step 1550,train loss = 0.0372,train accuracy = 99.22%,test accuracy = 88.28%

Step 1600,train loss = 0.0035,train accuracy = 100.00%,test accuracy = 89.06%

Step 1650,train loss = 0.0075,train accuracy = 100.00%,test accuracy = 87.50%

Step 1700,train loss = 0.0034,train accuracy = 100.00%,test accuracy = 87.50%

Step 1750,train loss = 0.0468,train accuracy = 96.88%,test accuracy = 91.41%

Step 1800,train loss = 0.0323,train accuracy = 97.66%,test accuracy = 86.72%

Step 1850,train loss = 0.0547,train accuracy = 96.09%,test accuracy = 84.38%

Step 1900,train loss = 0.0023,train accuracy = 100.00%,test accuracy = 93.75%

Step 1950,train loss = 0.0060,train accuracy = 100.00%,test accuracy = 92.85%

Step 2000,train loss = 0.0032,train accuracy = 100.00%,test accuracy = 91.89%