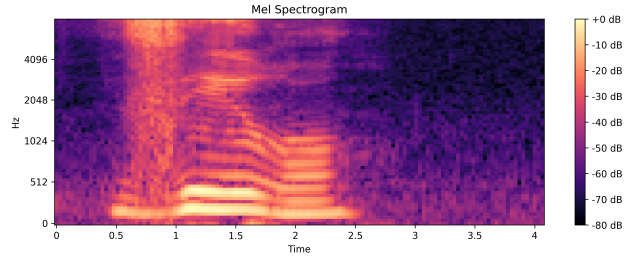
Advanced Topics in Audio Processing using Deep Learning
Assignment 2

Daniel Broyd Doron Aloni

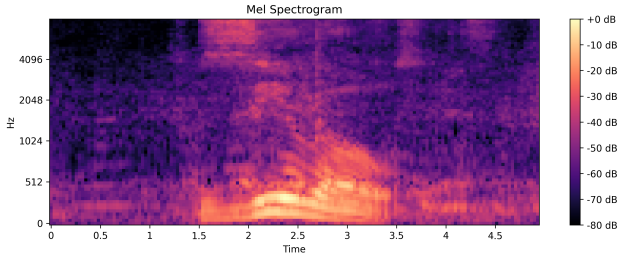
January 13, 2025

Data Acquisition

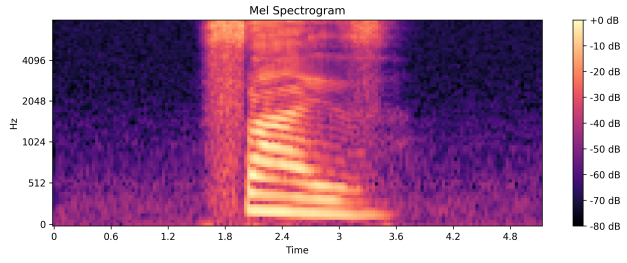
2. i. Differences within speaker samples:



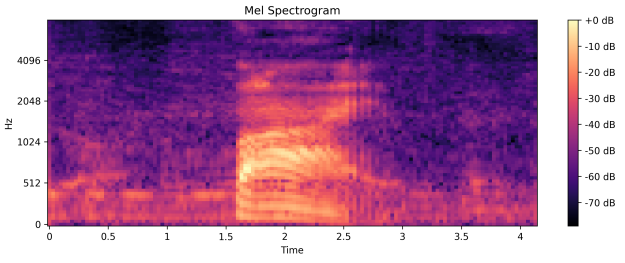
(a) Speaker 1 (Female): 0



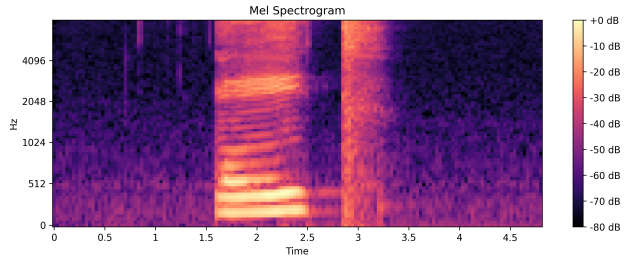
(b) Speaker 2 (Male): 0



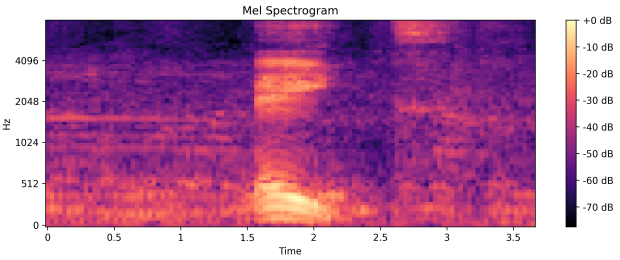
(c) Speaker 1 (Female): 5



(d) Speaker 2 (Male): 5



(e) Speaker 1 (Female): 8

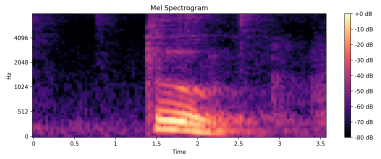


(f) Speaker 2 (Male): 8

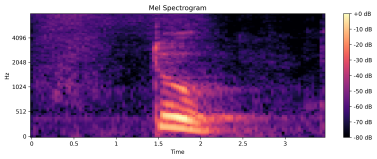
Figure 1: Differences within Speaker Samples

As can be clearly seen (across both speakers), different digits correspond to a different combination of phonemes.
Some digits dominate lower frequency bands (e.g. zero), while high frequency components are more prominents in other digits (e.g. eight).

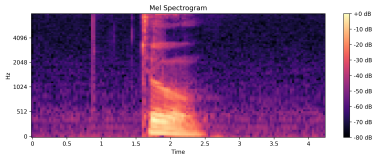
ii. Differences across Digit Samples:



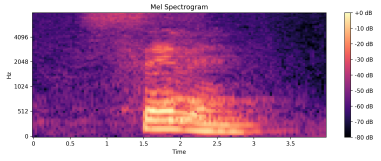
(a) Female, number 2



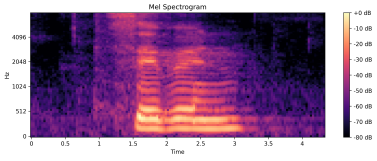
(b) Female, number 2



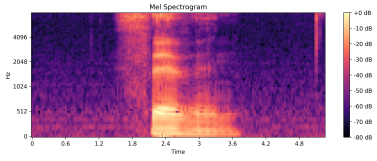
(c) male, number 2



(d) Female, number 7



(e) Female, number 7



(f) male, number 7

Figure 2: Overall caption

We can see that for a certain digit the mel spectrogram between different speakers is quite similar, though we can see that the male have much brighter yellow color and dominates the lower frequencies.

DTW

e. The distance matrix is:

```
[[15081.63 13997.51 11569.14 21567.7  14654.77 13963.56 16564.13 16311.44 15783.87 13655.17]
 [14068.53 15405.54 10914.14 13246.33 12659.07 13203.87 10486.57 11669.76 11365.88 10163.95]
 [12972.05 11768.56 21871.81 19363.54 15466.47 12050.39 26455.94 24896.13 16165.52 23681.58]
 [14082.23 16625.03 10189.41  9332.77 13249.16 14658.07  8741.31 11744.51 11990.57 11252.83]]
```

Figure 3: distance matrix

f. The accuracy over the training set is 0.2

g. The accuracy over the Validation set is 0.275 and the confusion matrix is:

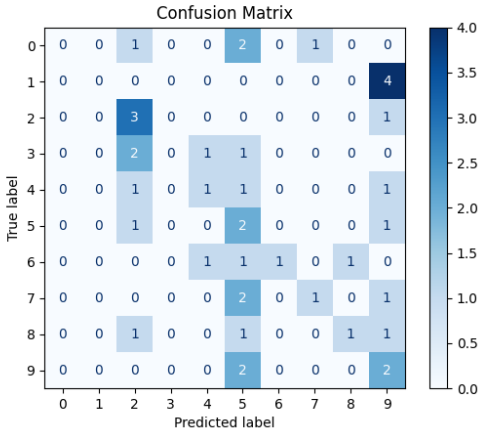


Figure 4: confusion matrix

h. The accuracy over the Validation set after normalizing is 0.3 and the confusion matrix is:

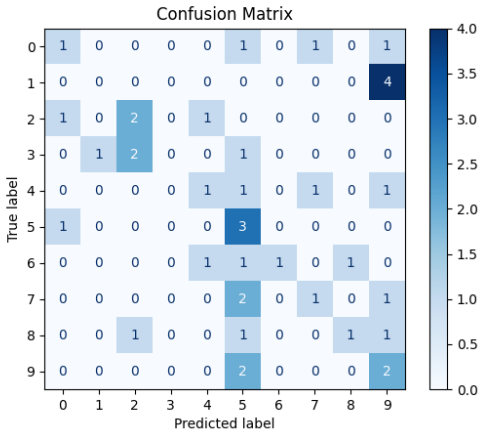


Figure 5: normalized confusion matrix

Forward Algorithm

5. c. The probability of the sequence "aba" is: 0.15

d. pred matrix is:

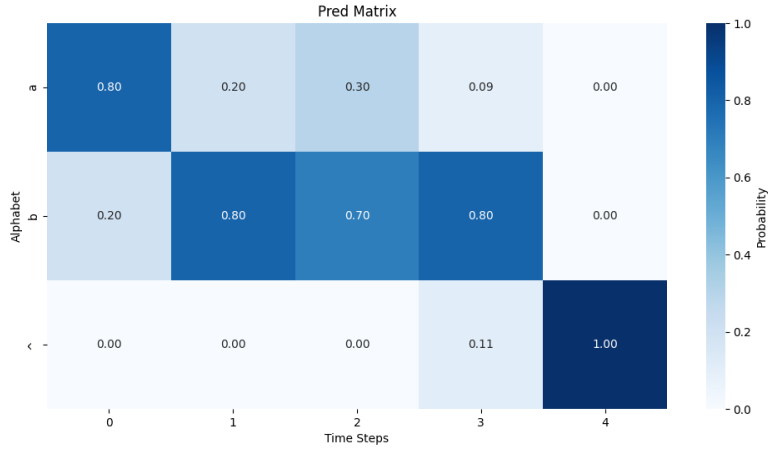


Figure 6: pred matrix with probabilities

6. The probability of the path is 0.00037, selected labels are: ”abbaˆ” and the path over the back trace matrix:

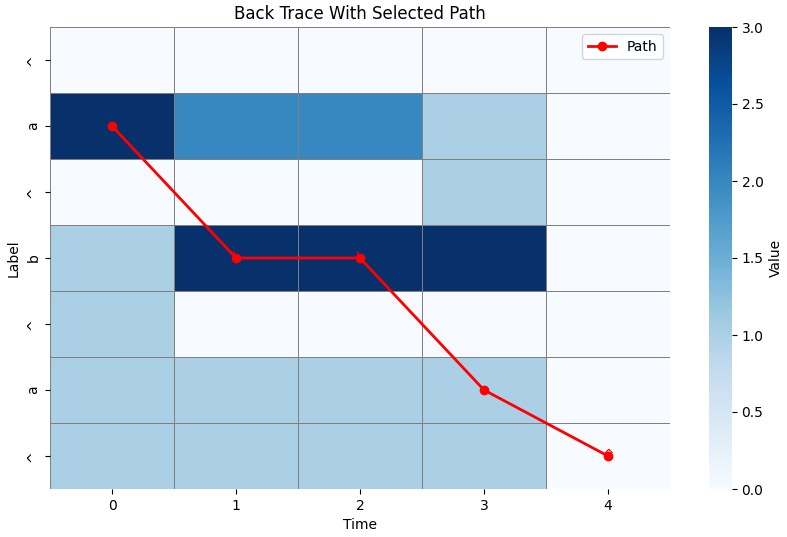


Figure 7: path on back trace matrix

7. The probability of the path is 0, this is because of numeric issues in python, it should be a very small number instead. The selected labels are
   ^^t ^^^h ^e ^n ^ ^g ^o ^o ^^d ^b ^y ^e ^^^^

   sa ^^i ^d ^ ^^r ^a ^t ^t ^^^s ^^^^^

   ^t ^h ^e ^y ^w ^a ^n ^t ^^^^^h ^o ^m ^e ^^^^^^^^^^^^^

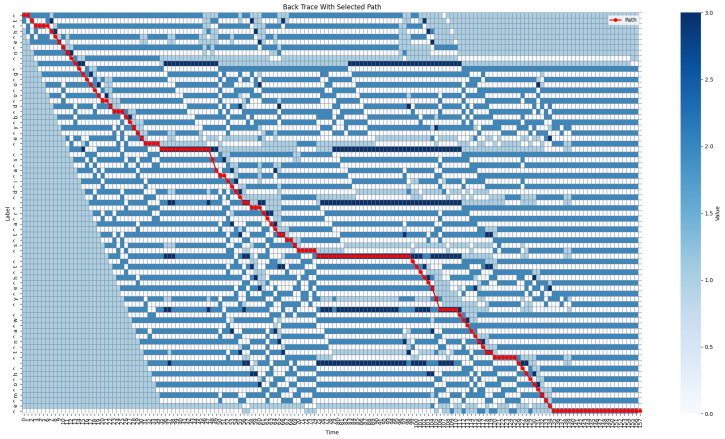


Figure 8: path on back trace matrix