Assignment 3

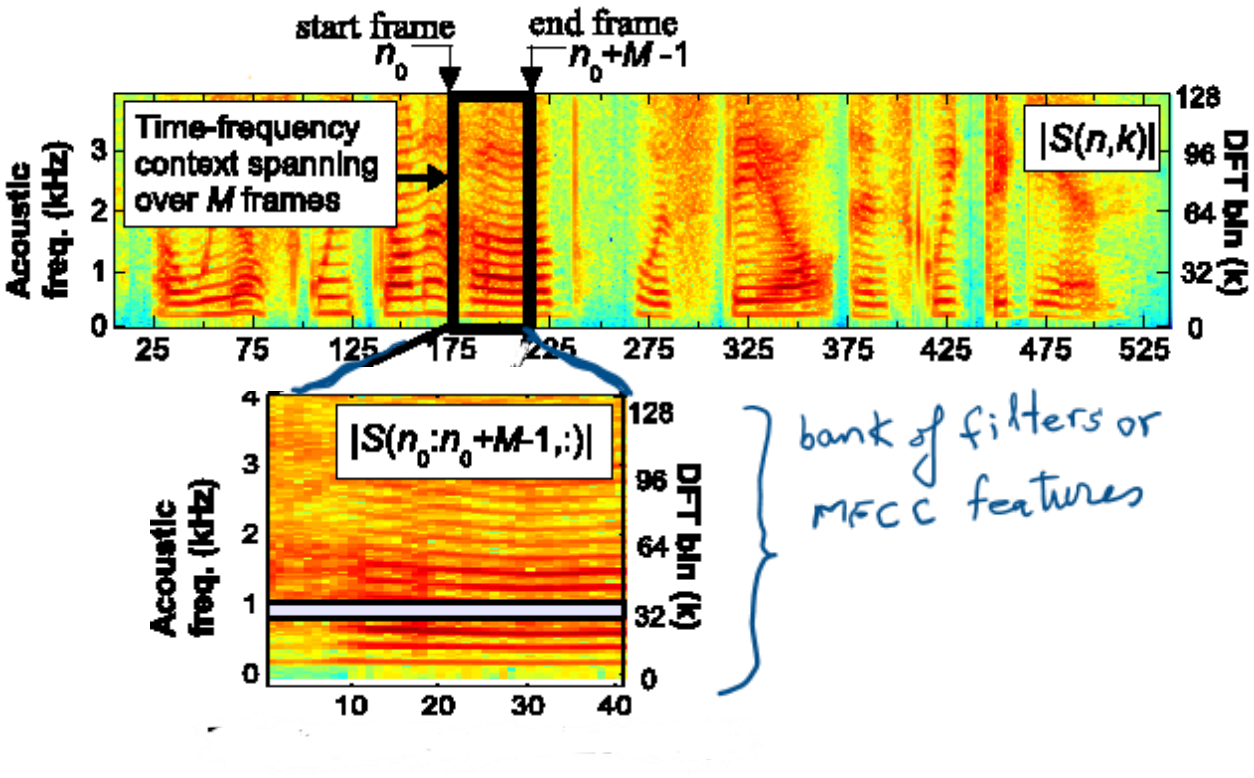
Part I

Problem 1: Data Augmentation and Data Synthesis

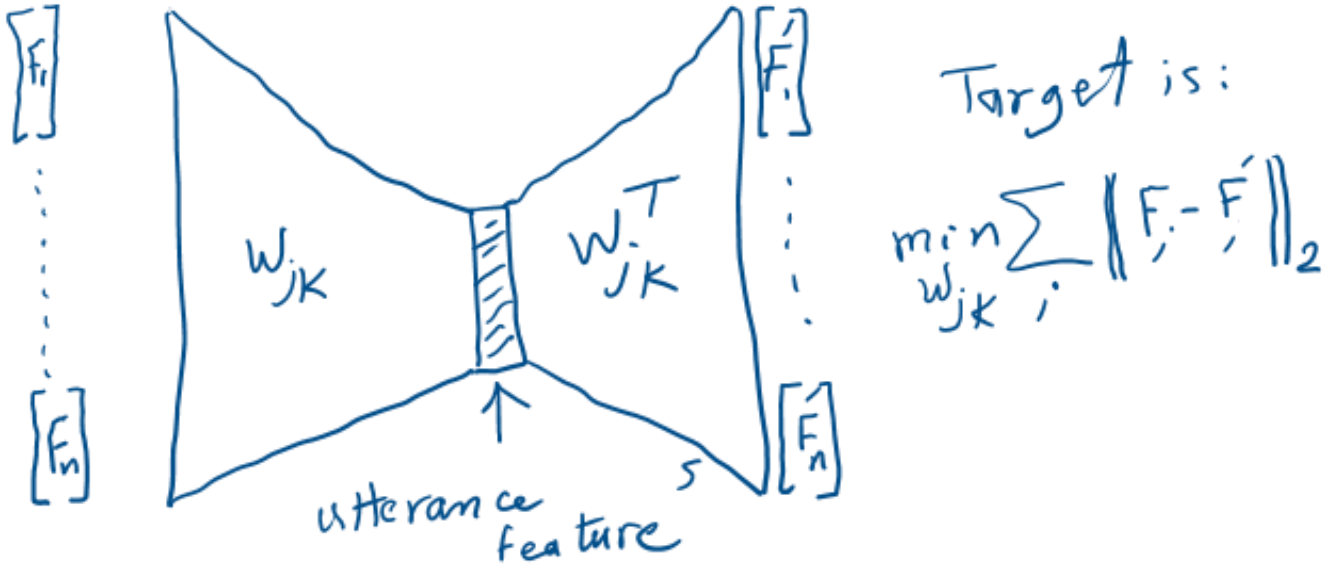
Autoencoder to represent and solve the utterances of the 10 digits given in ass 2:

Given the speech data (train and test) for the 10 digits uttered by many speakers, develop an autoencoder (AE) to generate a single vector that represents each utterance (you may divide the utterance into frames of 15 m. Sec. each, then concatenate these frames and use the AE to generate a single vector of a length of your choice to represent the complete utterance). Once you represent each utterance in a single vector, develop a classifier to classify each utterance in the test set.

1. You may start by just calculating the average frame for each utterance to get a vector that represents each utterance. Make it the baseline for your results.
2. You should try to use the AE in this way:
    a. Concatenate all the frames in one shot and generate a single vector for each utterance. However, you will have different lengths for each utterance, then use the maximum length and append zero frames in shorter utterances.
    b. Use the AE to generate a vector of the same length as the 1st and 2nd frames, then between the generated vector and the 3rd frame, and so on until you finish all the utterance frames.
    c. (optional with a bonus mark) Repeat 'b' above but test the combinations available between each consecutive two frames and start combining the lowest available error.



An example of a spectrum image for an utterance



An Auto Encoder is used to concatenate 2 or more frames to ultimately reduce the utterance to a single vector.

Problem 2: Data Augmentation

For the ReducedMNIST database:

1. Generate using augmentation methods different number of examples as indicated in the table below. Augmentation: include rotating right, left with different angles, moving the digit random movements by delta x and delta y, adding % of white noise…etc.
2. Design a pipeline to test the recognition model generated with the mixture shown in the table below.
3. Fill out this table of the results for the same test data (200 examples) and add your comments on the results.
   (Hint: we increase the data generated as no labeling is needed gust computer time to generate.)

		Real Data		
		300	**700**	**1,000**
No. of Generated examples	**0**	300 real	700 real	1000 real
	1,000	300 real + 1000 generated of 300 real	700 real + 1000 generated of 700 real	1000 real + 1000 generated of 1000 real
	2,000	300 real + 2000 generated	700 real + 1000 generated	1000 real + 1000 generated
	3,000	300 real + 3000 generated of 300 real	700 real + 3000 generated of 700 real	1000 real + 3000 generated of 1000 real

Problem 3: use GAN to generate Synthetic Data

1. Repeat the above work but the generated data in this case will be through using the GAN network to generate each digits the number of times shown in the table.
2. Use a table like the one shown above and generate a model in each case. Compare the results and comment.
3. Select any reasonable combination of the augmented and synthetic data in the case of 300 examples available from the real data and show to what extent we can reduce the need for real data. Compare the best case with the case of using all the available 1000 examples of real data.

Part II

effect of Attention

Problem 4: Understanding the Impact of Attention Mechanisms

a. Build a convolutional neural network (CNN) to classify images from the **ReducedMNIST** dataset. Then, build a second version of the CNN that includes a **spatial attention mechanism**. Compare the two models in terms of **accuracy** and **training time**. What difference does the attention mechanism make?

b. Revisit the spoken digits task from Assignment 2, where you used spectrogram images to recognize 10 spoken digits. Build two CNN models for this task: one **with an attention mechanism** and one **without** (which you implemented in ass-2). Compare their performance and explain how attention affects the accuracy and training time in this case.

For both parts (a and b), your submission should include:

- A clear and detailed report describing **(put your answers on a table)**:
    - The network architectures you used
    - Your training process and chosen hyperparameters
    - A comparison of performance between models with and without attention
    - Your analysis of how the attention mechanism affected the results
- Insights and observations you gained from your experiments.
- Suggestions for future improvements, such as trying different types of attention or tuning the model further.
  Bonus marks will be awarded for implementing and testing creative or improved ideas which must be stated clearly.