What needs to be written:

1. Introduce the work
2. Explain why it’s cool
3. How is it different
4. Good references

What need to be in the proposal:

1. Title
2. Name of advisor(s)
3. Group members (include class section)
4. Why doing this project?
5. How it impact current research we are doing
6. Describe how the project will be partitioned into individual contributions
7. Show brief description of the data
8. Show preliminary results/Data processing
9. What are the classification techniques
10. Make sure to state what is the “ground truth” (a point where you can compare the performance of our algorithms)
11. Acknowledge that there will be written research paper submitted prior to the final exam week and there will be a short oral/poster presentation during the final week.
12. **Inspiration**

Recent efforts in applications such as image recognition has been fundamental in creating the ground for further endeavors in other fields. Therefore, while image classification is mostly solved, with more than 98% accuracy in some dataset, speech classification still has a lot of room for improvement. Hence, it is worth exploring and improving existing learning techniques in this research.

The performance of each method will be evaluated with respect to a simple K-nearest neighbor classifier.

**Gender Classification Through Speech Research Proposal**

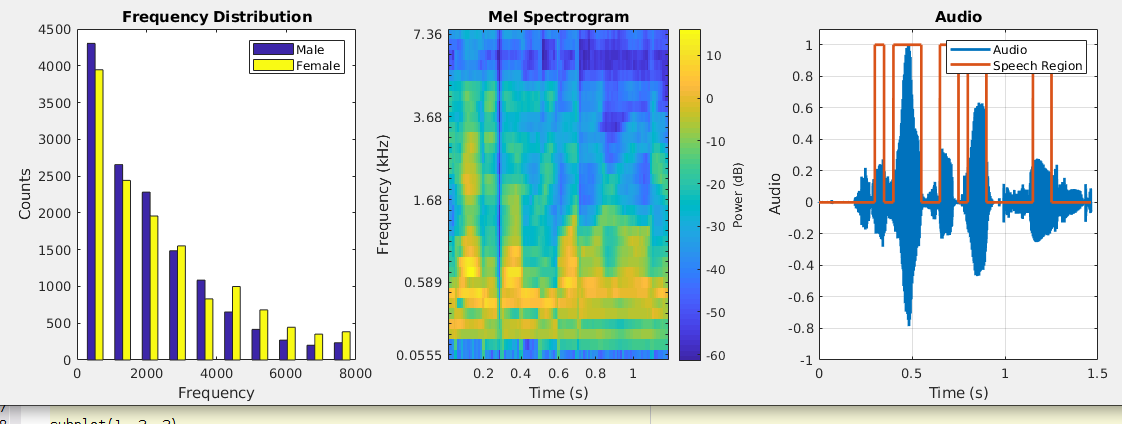
*Authors: Hieu Mai, Tai Nguyen, Harpreet Cheema - ECEC 487 - Advisor: Dr. Anup K. Das*

1. **Introduction**

Speech recognition is a complex problem with real-life applications. Hence, this research focuses on exploring different learning techniques: Convolutional Neural Network (Hieu Mai), Recurrent Neural Network (Harpreet Cheema), and Spiking Neural Network (Tai Nguyen), in gender classification and speech data processing.

1. **Preliminary Data Processing**

The dataset consists of 6300 sentences, 10 sentences spoken by each of 630 speakers (genders specified) from 8 major dialects in the US.



*Figure 1 (left): Dominant audio frequency distribution after silence removal of the male and female population. Figure 2 (center): Mel Spectrogram of a sample audio, Figure 3 (right): example of the silence removal algorithm*

1. **Methodologies**

The SNN algorithm in this research is a feedforward network using spike-timing-dependent plasticity (STDP) learning rule. The network consists of an input layer, a convolutional layer, and a pooling layer, whose output determines the class of the data [4]. To obtain the input signal for the SNN, the Mel-Frequency Cepstral Coefficients (MFCCs) are obtained from the raw data with a frame size of 50 ms with 25 ms overlap [4], which is encoded with the time-to-first-spike method. This method will be tested against another method, in which the input signal is the 10 most dominant frequencies for each recording.

1. **References**

[4] Dong M, Huang X, Xu B (2018) Unsupervised speech recognition through spike-timing-dependent plasticity in a convolutional spiking neural network. PLoS ONE 13(11): e0204596. <https://doi.org/10.1371/journal.pone.0204596>

Speech recognition with CNN: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/CNN_ASLPTrans2-14.pdf>

Gender classification CNN: <https://publications.idiap.ch/downloads/papers/2018/Kabil_INTERSPEECH_2018.pdf>

RNN

<https://pdfs.semanticscholar.org/536c/3a308763eaa5ce44e2e35c8efb98bf341fd9.pdf?_ga=2.11260471.28179894.1571696644-1697620773.1571418354>