Link to Zoom Meeting: https://utoronto.zoom.us/j/3813974797
Pass: 095799

Meeting Nov 21 12pm

Plan:
Ken: Get MFCC done, hopefully trimmings as well.
https://towardsdatascience.com/extract-features-of-music-75a3f9bc265d

Daniel & Wendy: Think about how to create the first layer with MFCC as inputs

MFCC
The mel-frequency cepstrum (MFC) is a method of representing the power spectrum of a sound. The MFCCs, mel frequency cepstral coefficients, are the values that make up the MFC which are equally spaced on the mel scale. This representation allows us to represent how a human's auditory systems process sound and is often used in sound processing related problems [1].

One of the major features of an audio signal is the frequency of the signal, often referred to as the pitch. However, a human's perceived frequency for a signal often does not line up with the actual frequency. Humans can detect ranges of frequency from 20 Hz to 20 kHz, but our ability to detect changes in frequency are better for lower frequencies. For example, while the distance between a 300 Hz and 400 Hz signal and a 900 Hz and 1 kHz signal are identical (100 Hz), we may perceive a greater difference between the 900 and 1000 Hz signal. As a result, the mel scale was created, which relates perceived to actual frequencies and was experimentally determined. This scale takes into account that humans ear as a filter that concentrates more on lower frequencies than higher ones [2].

Extra reading:
http://www.speech.cs.cmu.edu/15-492/slides/03_mfcc.pdf

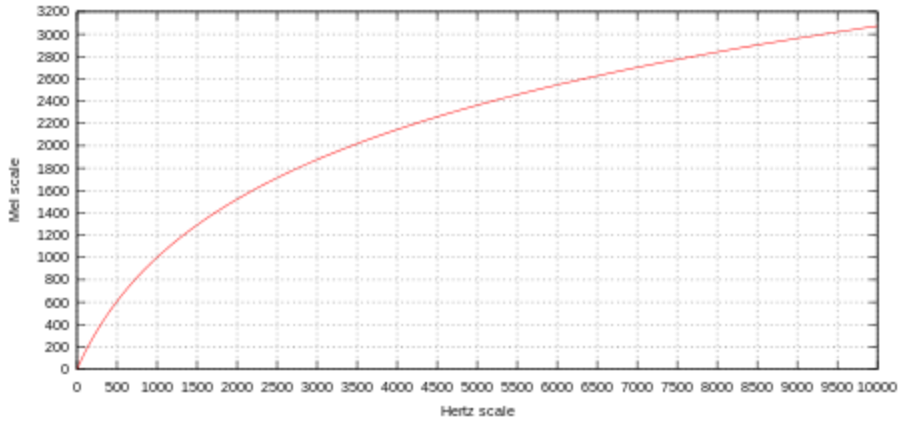
[1] https://en.wikipedia.org/wiki/Mel-frequency_cepstrum
[2] https://medium.com/prathena/the-dummys-guide-to-mfcc-aceab2450fd

Mel Scale

The mel scale shows the relationship between perceptual and actual pitches for humans, and provides a way to measure frequency based on pitch difference.
The common formula for this transformation is given by

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

With the corresponding graph



https://en.wikipedia.org/wiki/Mel_scale

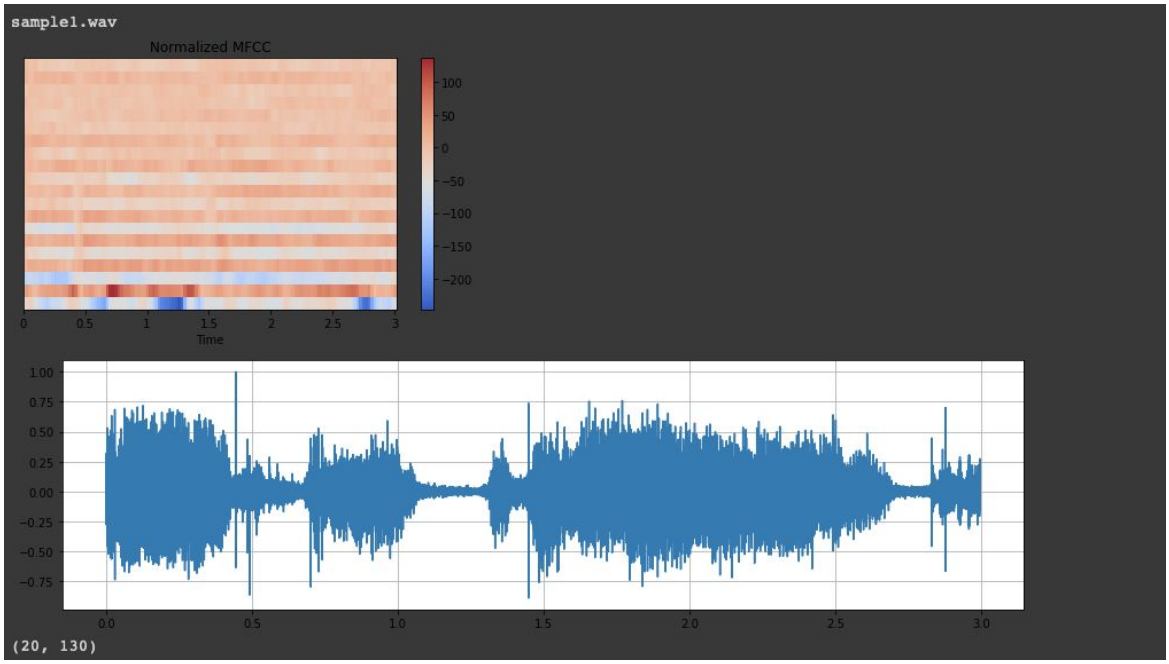
Good explanation of MFCC and use of DCT
https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC

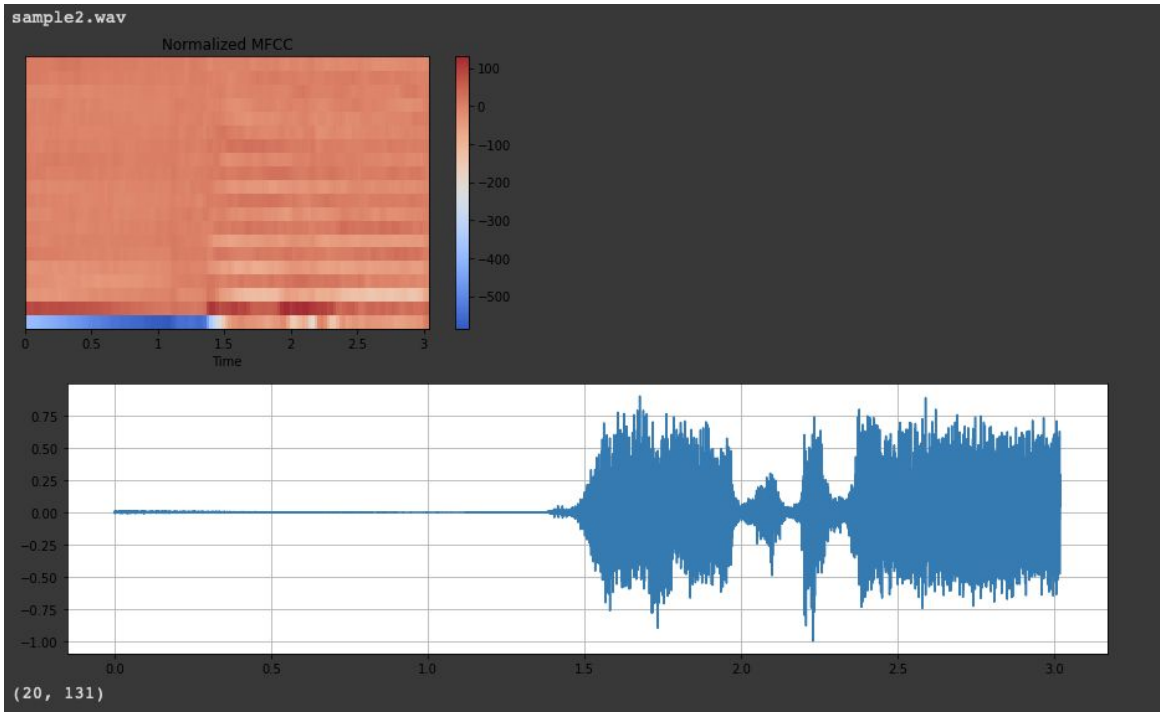
SVM and use of MFCCs
https://www.irjet.net/archives/V5/i9/IRJET-V5I9170.pdf

Meeting Nov 23 12pm

https://colab.research.google.com/drive/10EtzeljO1qBOYft5eAZpVyOLpBpg82Vh?usp=sharing

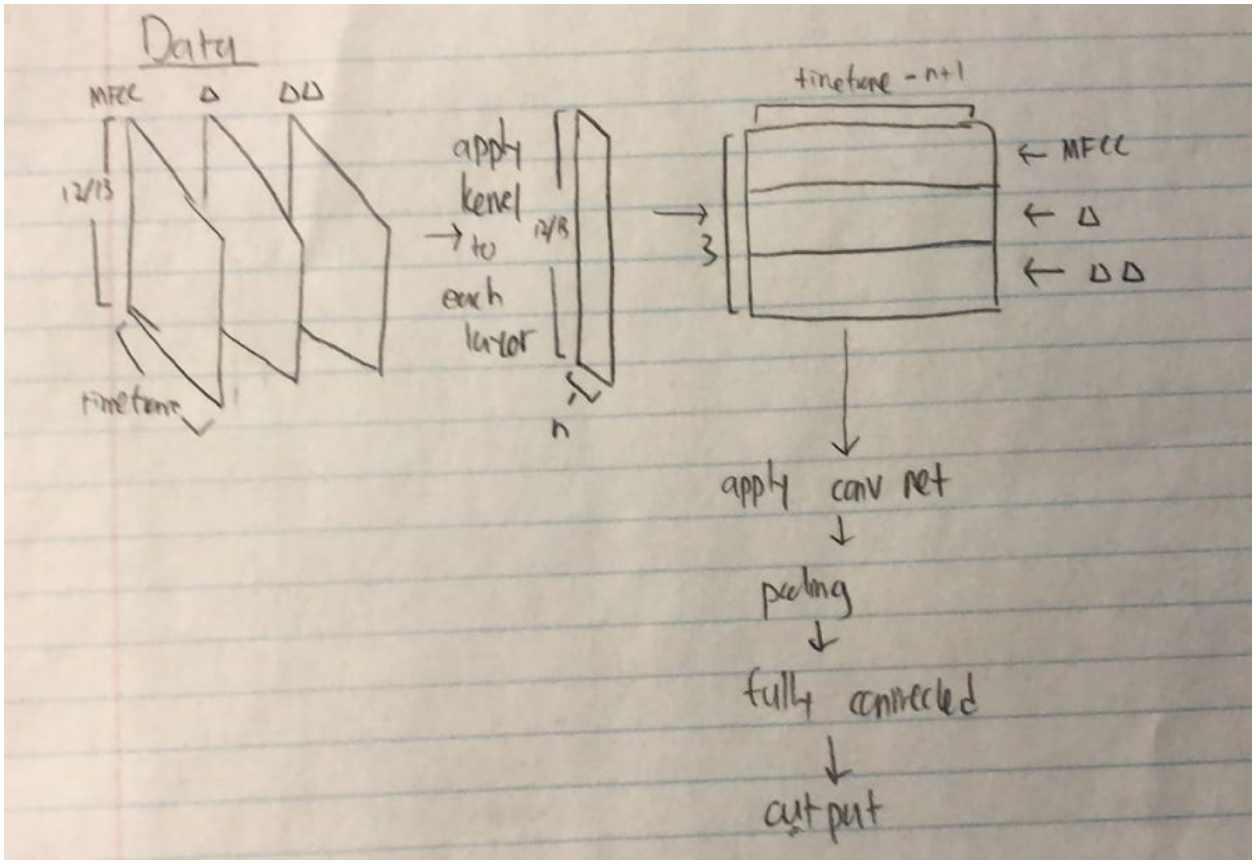
Plan for next meeting:
Ken: Create 1000, 3 second files: 500 lions and 500 hyenas
Wendy: find ‘bad’ sound files





Meeting Nov 24 9pm

Network Architecture



1. Use 3 convolutional networks for the kernel data transformation (1 for each one).
2. Stack the transformed layers on top of each other
3. Use 2 convolutional layers to extract features
4. Max pooling
5. Fully connected

Note: steps 3-5 are the main training part where we play with different hyperparameters

Daniel: Complete step 1 and 2

0 = hyenas
1 = lions

Train test split: 75-25?
Hyperparameters

- Loss function
    - CE
    - MSE

- Optimizer
    - SGD
    - Adam
- Kernel Size
    - 3x3
    - 1x1
    - 5x5
    - 7x7
- Number of fully connected Layer
    - 1
    - 2
- Number of convolutional layer
    - 5
    - 1
    - 2
- Activation Function
    - ReLU
    - Sigmoid
    - softmax
- Number of kernels on each convolutional layers
    - 64
    - 32
- Learning Rate
- Batch Size
- Epoch Size
    - 10
    - 25
    - 50
- Max pool