Deep Sample

A Study of Audio Segmentation

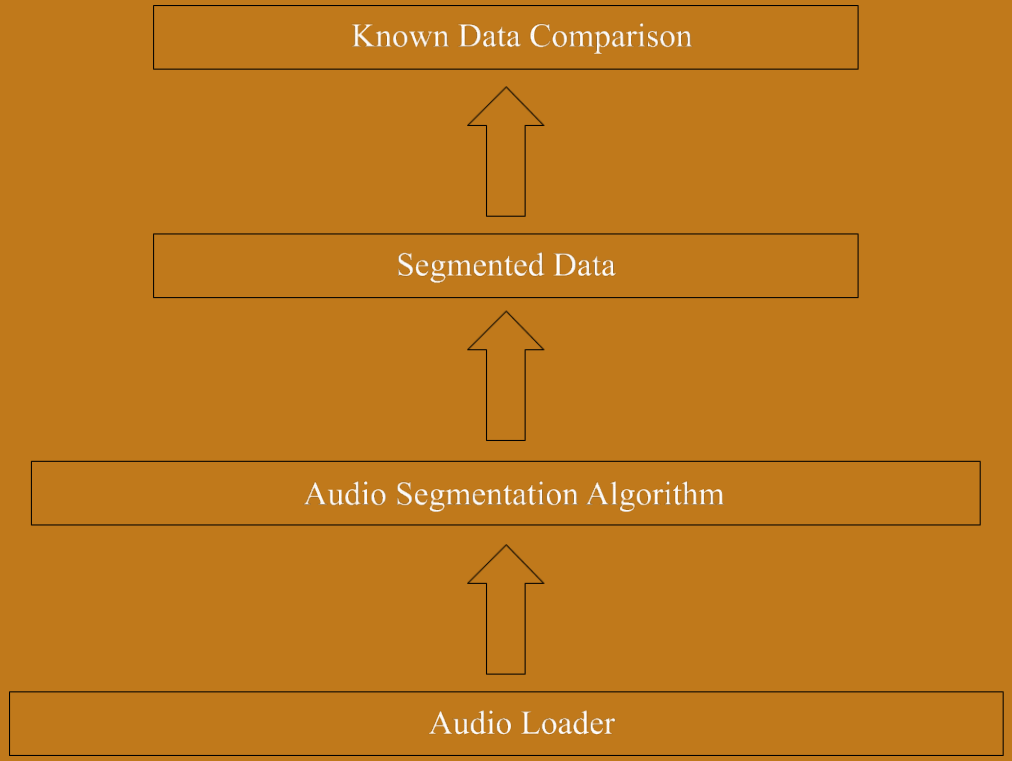
Andrew Moore, Hue Truong, Alex Reno
Quinsigamond Community College

Background

What is Audio Segmentation?

- Preprocessing step in signal analysis
  - Breaks signal into component parts
- Modern practices use multiple segmentation algorithms, often paired with machine learning techniques.
- Deep Sample focuses on the Zero Crossing, Spectrum Centroid, Spectrum Flux, and Real Cepstrum algorithm.

Audio Segmentation Pipeline



Fourier Transform

- Decomposes a signal into its component frequencies
- Derived from the fourier series, where the limit of its periodicity approaches infinity
- Useful for looking at the frequency of a signal

$$F(x) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx$$

Real Cepstrum

- Transformation of the Fourier Transform
- Defines real parts of the signal
- Used to analyze amplitude of a spectrum over time

$$w(n) = 0.54 - 0.46 cos(2\pi n/(M-1))$$

Spectrum Centroid

Center of gravity of the spectral distribution, or the weighted average of frequencies

$$C_t = \frac{\sum_{n=1}^{N} M_t[n] * n}{\sum_{n=1}^{N} M_t[n]}$$

Spectrum Flux and Zero Crossing

Spectral Flux is the measure of the average signal change across an audio wave.

The Zero Cross is a measure of the noisiness of the signal.

$$F_t = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2$$

$$Z_t = \frac{1}{2} \sum_{n=1}^{N} |sign(x[n]) - sign(x[n-1])|$$

Methods

Fast Fourier Transform And Real Cepstrum

- Fast Fourier Transform:
  - Precursor for zero-crossing and spectral algorithms
  - Implemented using direct model of Fourier equation

$$F(x) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \xi} dx$$

- Real Cepstrum:
  - Raw signal passes through Fourier Transform
  - Passed through hamming window to taper off data at each end
  - Loops through each vector element to take its logarithm
  - Passed through an inverse Fourier algorithm

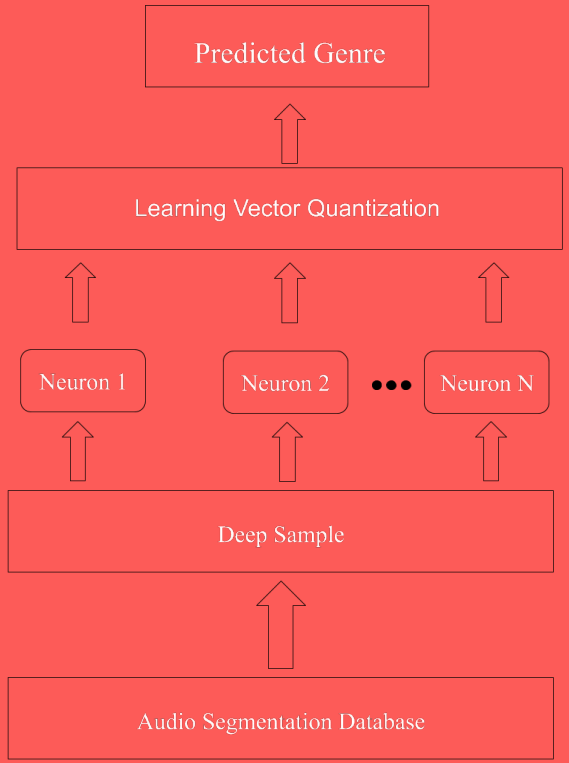
Spectrum Centroid

- Spectrum Centroid runs in a single loop:
  - Find the FFT if not already generated
  - Loop over each channel:
    - Calculate the numerator term
    - Sum the numerator terms
    - Sum the denominator terms
  - Take the quotient
- O(n) runtime, bounded by the size of the Fourier transform

Spectrum Flux and Zero Crossing

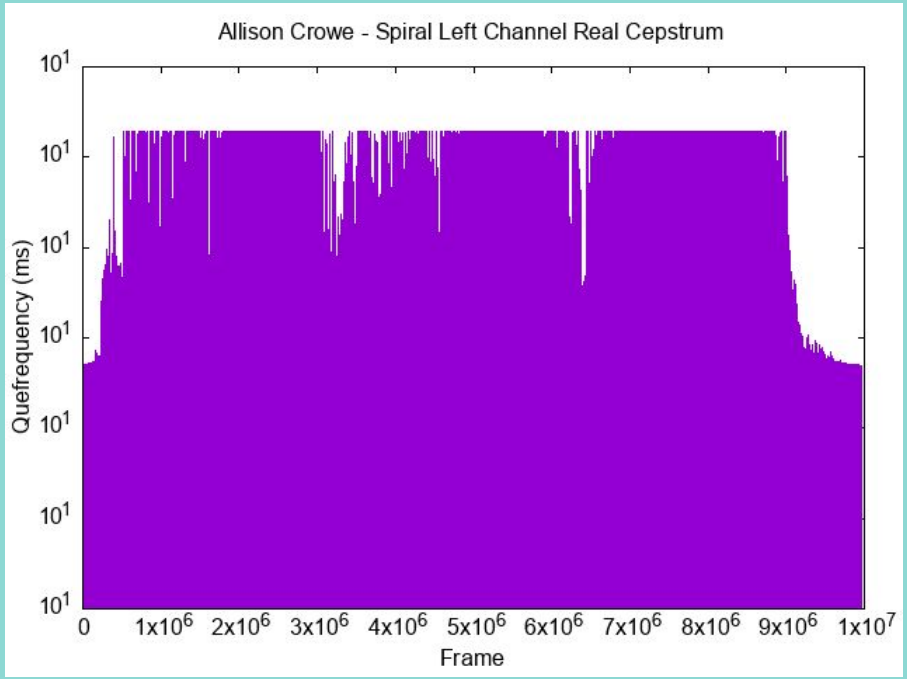
- Spectrum Flux is implemented as follows:
  - Loop through each channel
  - Calculate normals
  - Sum the differences squared.
  - O(n) runtime
    - Bounded by sample set size.
- The zero cross is runs over each channel in a nested loop
  - Calculates the sign of each element and sums the differences
  - Takes ½ the sum as the zero cross value of that frame.
  - $O(n^2)$ runtime
    - Bounded by the square of the sample set size.

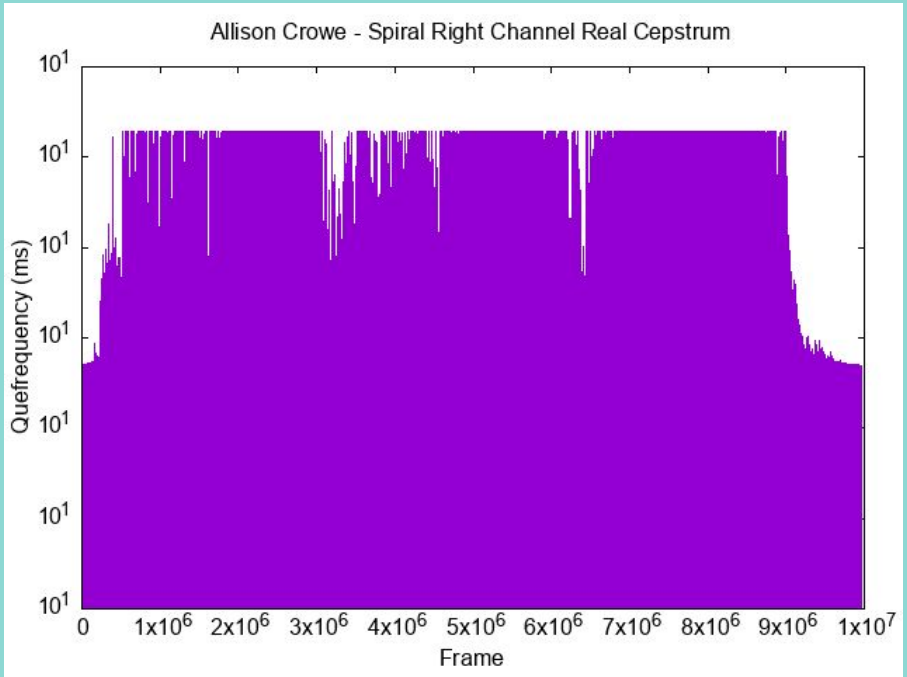
Deep Sample ANN



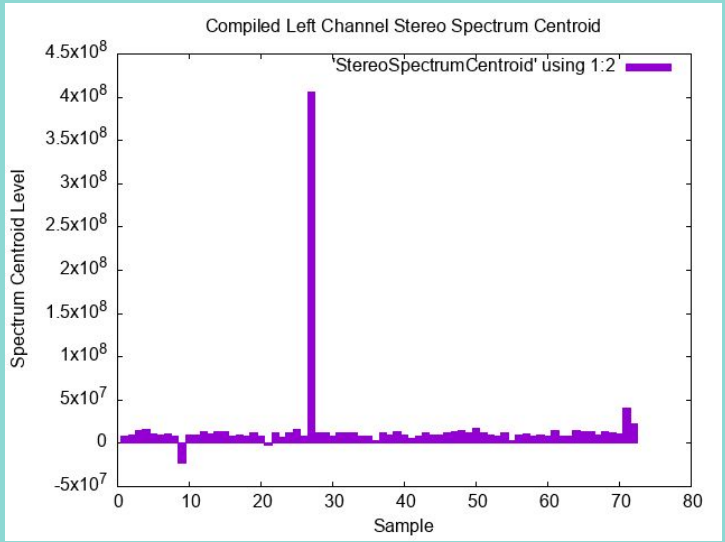
Results

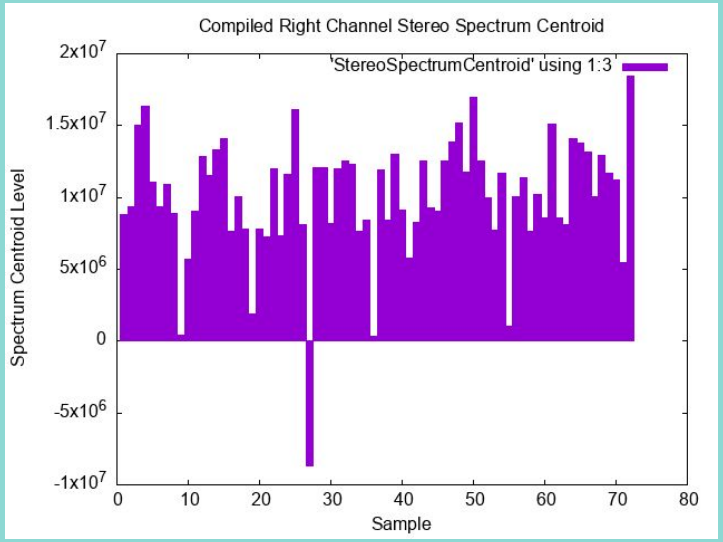
Real Cepstrum





Spectrum Centroid





Spectrum Flux

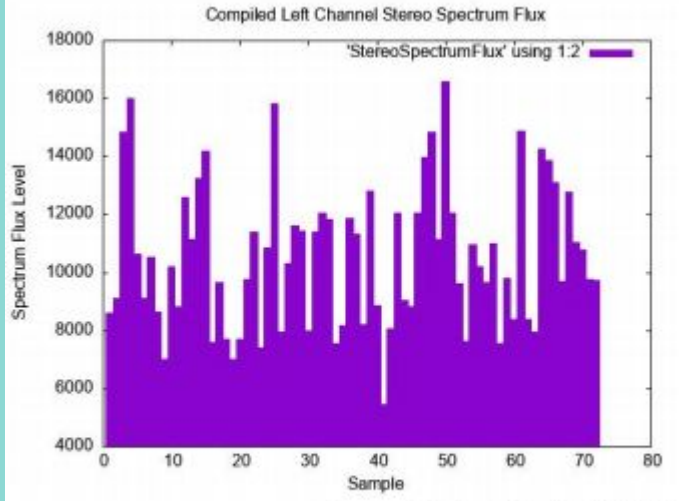


Figure 11: FreeMusic2011 [7]
Compiled Left Channel Spectral Fluxes

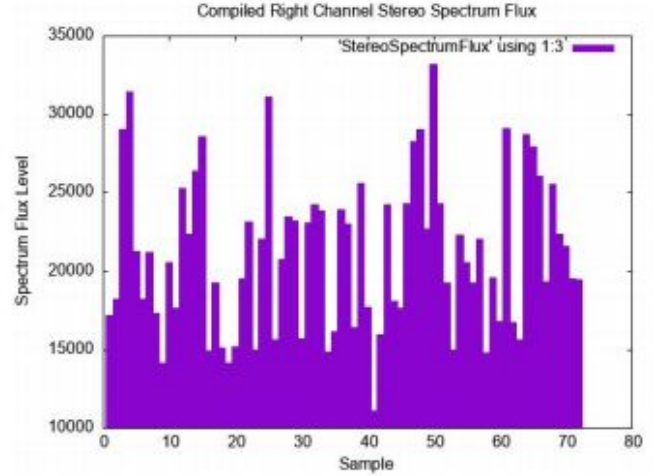
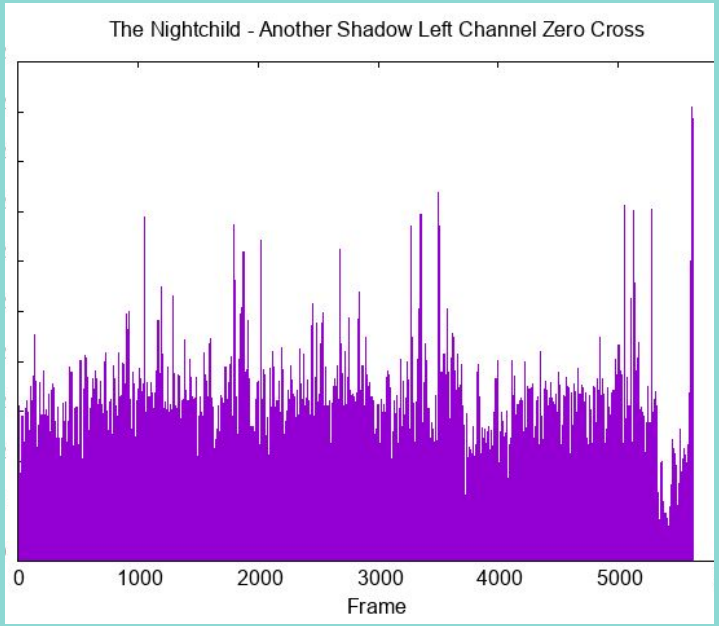
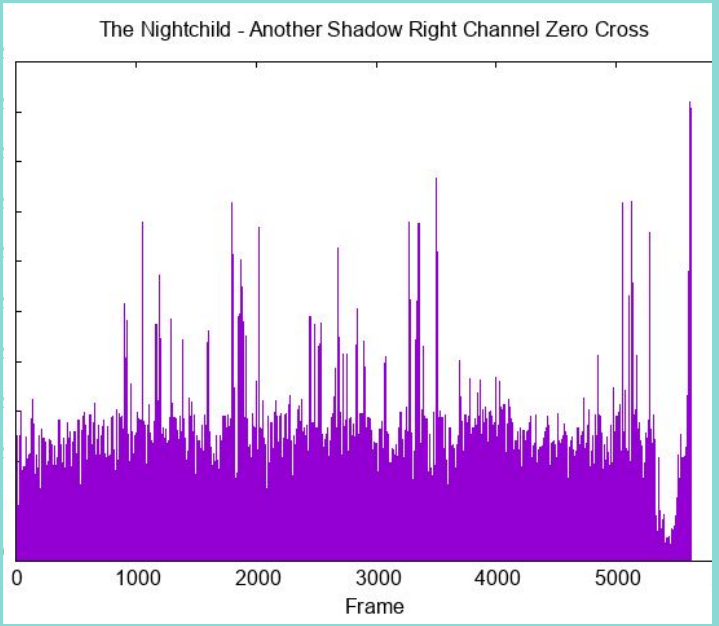


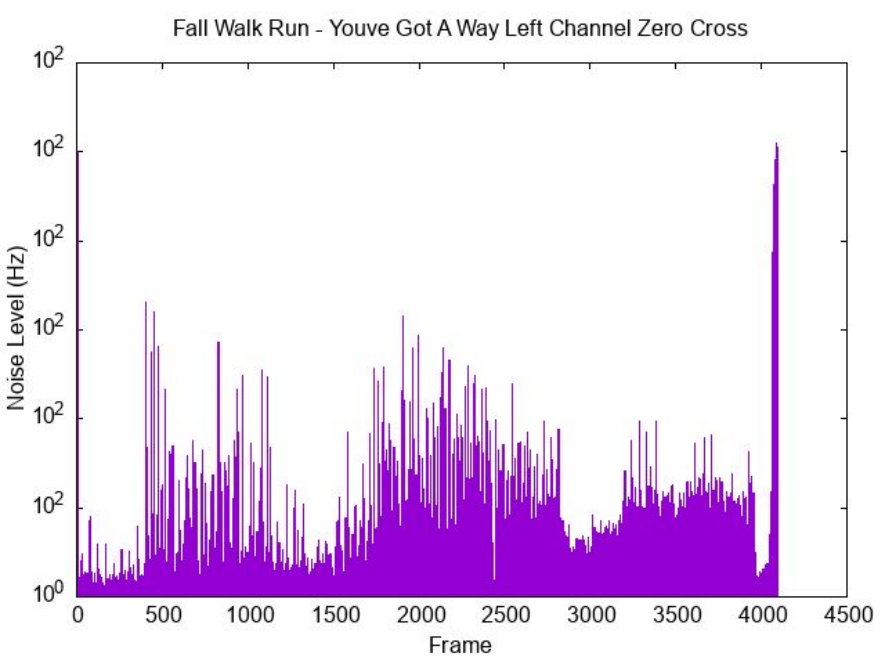
Figure 12: FreeMusic 2011 [7]
Compiled Right Channel Spectral Fluxes

Zero Crossing

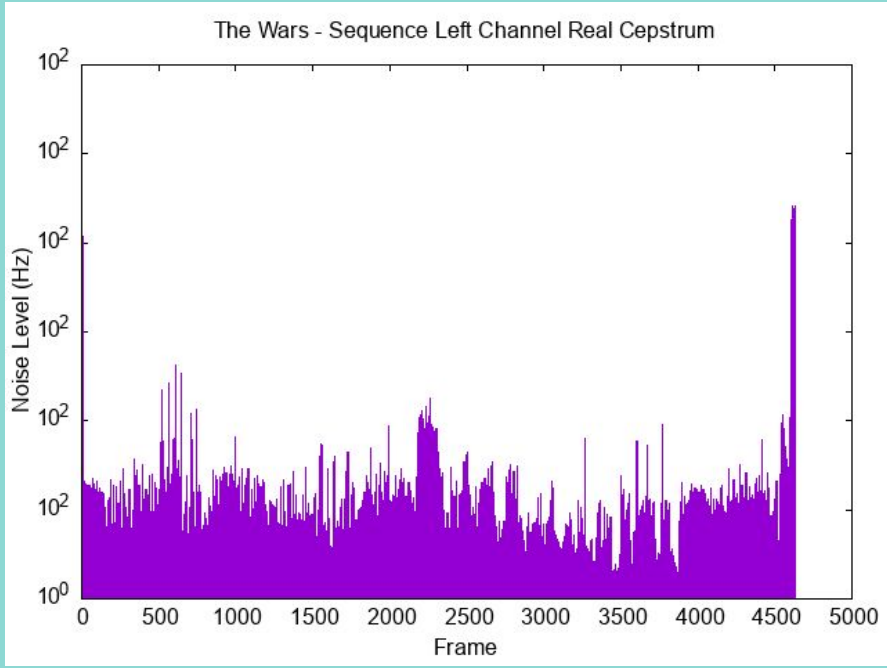




Deep Sample

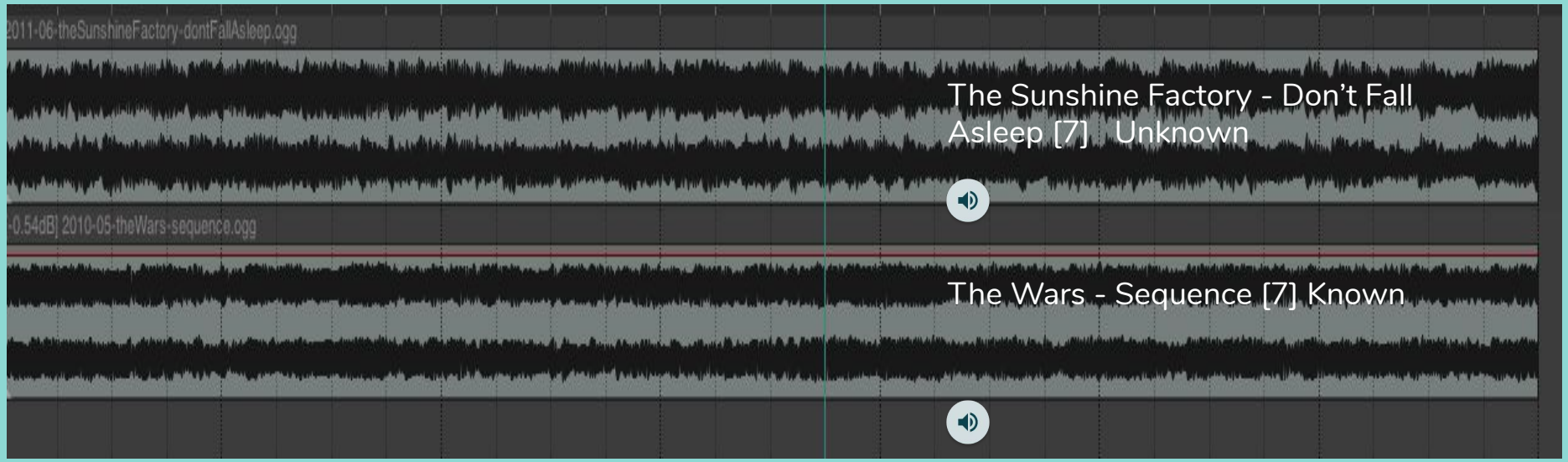


Unknown Sample Being Analyzed [7]



Proposed Match From Database [7]

Deep Sample Match



Combined Samples:

Future Work

- Move to mixed implementation with Matlab
- Expand the training set
- Enable DeepSample to learn
- Add additional segmentation algorithms
- Improve existing algorithms
- Improve data transfer methods or data compression methods

Works Cited

[1] Zahid, S., Hussain F., Rahid M., Yousaf M.H., & Habib H.A. (2015). “Optimized Audio Classification and Segmentation Algorithm by Using Ensemble Methods”. *Mathematical Problems in Engineering*, 2015, 1-11.
doi: 10.1155/2015/209814

[2] Steeb, W.H., Steeb W.H. (2005). “Mathematical Tools in Signal Processing with C and Java Simulations”. Hackensack, NJ: World Scientific.
doi: 3-8-2020

[3] Giannakapoulos, T., Pikrakis, A. (2014). “Spectral Flux”. Introduction to Audio Analysis.
https://www.sciencedirect.com/topics/engineering/spectral-flux

[4] Chu, W.T. (2014). “Musical Genre Classification”. Multimedia Content Analysis.
https://www.cs.ccu.edu.tw/~wtchu/courses/2014f_MCA/Lectures/Lecture%209%20Audio%20and%20Music%20Analysis%202.pdf

[5] Tzanetakis, G., Cook, P. (2002). “Musical genre classification of audio signals”. IEEE Trans. On Speech and Audio Processing, vol 10, no. 5, pp. 293-302

[6] Brownlee, J. (2016 Nov 14). How to Implement Learning Vector Quantization (LVQ) From Scratch with Python. machinelearningmastery. https://machinelearningmastery.com/implement-learning-vector-quantization-scratch-python/

[7] Multiple. (2009 Apr 26). Free Music Charts. https://archive.org/details/free-music-charts?tab=collection

Questions?