Learning Musical Instruments through Spectral and Temporal Analysis

Daniel D. Zundel
CprE585x - Developmental Robotics
24th April 2011

Abstract

Developmental robotics seeks to create programs that can become aware of themselves and their place in their environment. By studying how humans and animals develop awareness, interact with objects, and explore their environment, algorithms can be developed that can give programs or robots the ability to develop skills and abilities using similar methods. Humans learn to distinguish sounds through characteristics like pitch, rhythm, and loudness. This paper is to explore the feasibility of creating a program that has the ability to learn and eventually develop the means to discern between musical instrument families through spectral and temporal classifiers.

Table of Contents

Introduction

We seem to be able to differentiate sounds very easily. It takes little effort to understand loud and soft sounds, high pitches and low pitches, even the more complex characteristics like rhythms and timbre. We can listen to music and identify a piano, or guitar, or trumpet. In only a few notes we can discern what instrument is playing the sound. Why is it hard then, if we can do it so easily, to design a program that can do this as well?

What is it that makes one instrument different from another and gives it its own unique sound? It's called timbre. It's a very complex component of sound. Unlike pitch or loudness it cannot be mapped to a single scale. Some ways timbre is described is tone quality or color. When describing a sound one may say, "It's a dark tone", "A bright tone" or may use descriptors like warm, rich, smooth, rough, etc... This is where an instrument gets its identity. It's what separates a trumpet from a trombone or a clarinet from a flute. So how do we create a definition for something so complex? We can look into what can affect it.

Fletcher (1934) "...timbre depends principally upon the overtone structure; but large changes in the intensity and the frequency also produce changes in timbre." [1]

Looking into the spectral and temporal properties of sounds we may find an answer to how we can differentiate the timbre of one instrument with that of another mathematically. This will give us the tools needed to design a program or robot with the ability to distinguish between the different instrument families.

There are many applications that a program or robot with this ability could be used for including automatic music annotation, database retrieval systems, and why not even a conductor. Commercially there is more interest in speech recognition but this approach is another step in exploring how a robot or program can interact with sound.

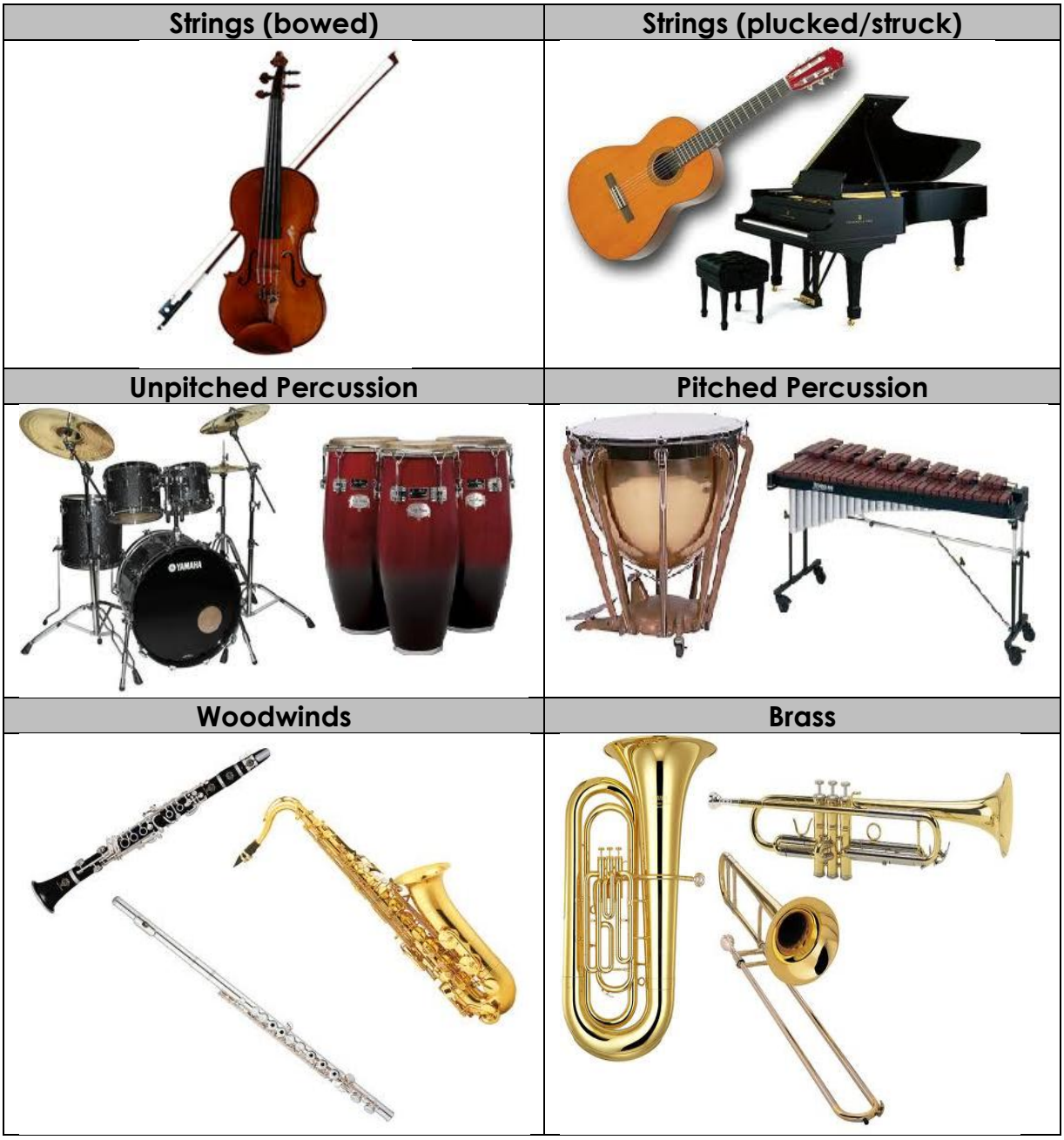

Figure 1 - Instrument Family Classification List

The purpose of this experiment is to develop an algorithm that will give a program or robot the ability to detect it has heard a sound, analyze that sound, and then determine if that sound belongs to family it has heard already or if it is different enough to be called a new family of instruments.

Approach

Almost all music we listen two is polyphonic. Multiple instruments playing together and playing different notes and phrases. This adds to the complexity of designing a robot or program that can identify which instruments are being played. Trying to perform a spectral analysis on several notes of different frequencies becomes quite an intensive task and for this reason a monophonic approach was taken.

A variety of instruments will be used in this experiment one at a time and use single isolated tones. An attempt was made to gather recordings from a local university with a variety of instruments and a wide range of pitches to get a high fidelity recording. Unfortunately this appointment fell through so instead a library of instrument recordings was gathered from the philharmonia orchestra website. [2]

In order to analyze each recording certain features of the notes will have to be extracted in both the time and frequency domains. In order to implement pattern recognition systems it is imperative that we have good feature extraction methods that are accurate and consistent.

Classifiers for instrument identification

Musical instrument signals usually consist of two parts. The first part is called the transient which occurs near the beginning of the sound in part that is called the attack of the signal. The second part is called the steady state or also the quasi-periodic part which occurs after the attack.

The attack time is an important feature. It is defined as the time it takes to reach the maximum amplitude of a signal from a minimum threshold magnitude.

$$attack\ time = t_{max} - t_{threshold}$$

Another feature is the amplitude envelope which describes the energy change of the signal in the time domain. From the RMS we can determine our attack time and our steady-state. One widely used method for computing the amplitude envelope is the Root Mean Square (RMS) method:

$$RMS = \sqrt{\left(\frac{1}{L}\sum_{n=0}^{L-1} x[n]^2\right)}$$

L is the length of the window and n is the sample number. A reasonable window length and window step size must be chosen. A large window tends to smooth out the envelope losing a lot of the transient effects and a very small window will introduce too much transient information.

The last feature is the spectral centroid which is commonly associated with the measure of the brightness of a sound and is defined in the following equation:

$$Centroid = \frac{\sum_{n=0}^{N-1} f(n)\,x(n)}{\sum_{n=0}^{N-1} x(n)}$$

It also is called the center of mass for the frequency components of a sound. Giving the power distribution over frequency.

By extracting these features of sound from each instrument certain patterns are expected to begin to develop which will be used in the analysis methods to determine if that sound is a unique family or one already categorized.

Analysis Methods

There are a variety of methods that have been used for speech analysis and recognition as well as sound recognition. They include the k-Nearest Neighbor algorithm, Naive Bayesian Classifiers, Decision Trees, Discriminant Analysis, Higher Order Statistics, Artificial Neural Networks, etc...

Two methods were looked at as possible choices for analyzing the extracted feature sets. The first method choice was to use a binary decision tree. Each node has two resulting branches. An example for the first decision node would be to compare attack times. An attack time less than 20 ms, like plucked strings for example, would branch off to one node while anything greater would branch off to another as shown in Figure 2 - Binary Decision Tree.

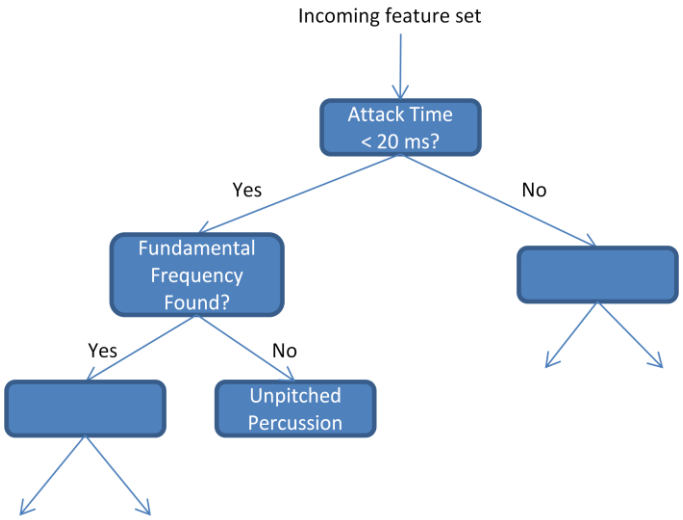

Figure 2 - Binary Decision Tree

Some drawbacks would be that the tree can become very complex very quickly but it gives good insight in viewing the choices made at each node which can help us to improve the classification decisions.

The second method choice was the k-Nearest Neighbor algorithm (k-NN). The main reason for this choice would be ease of implementation. Several "learned" data sets would be in memory to compare incoming sets against. If the choice of features have enough significant differences then a high accuracy of identification should be achievable. This has a hierarchical system similar to the binary tree but has a greater number of branches from each node as shown in Figure 3 - k-Nearest Neighbor Classification Hierarchy.

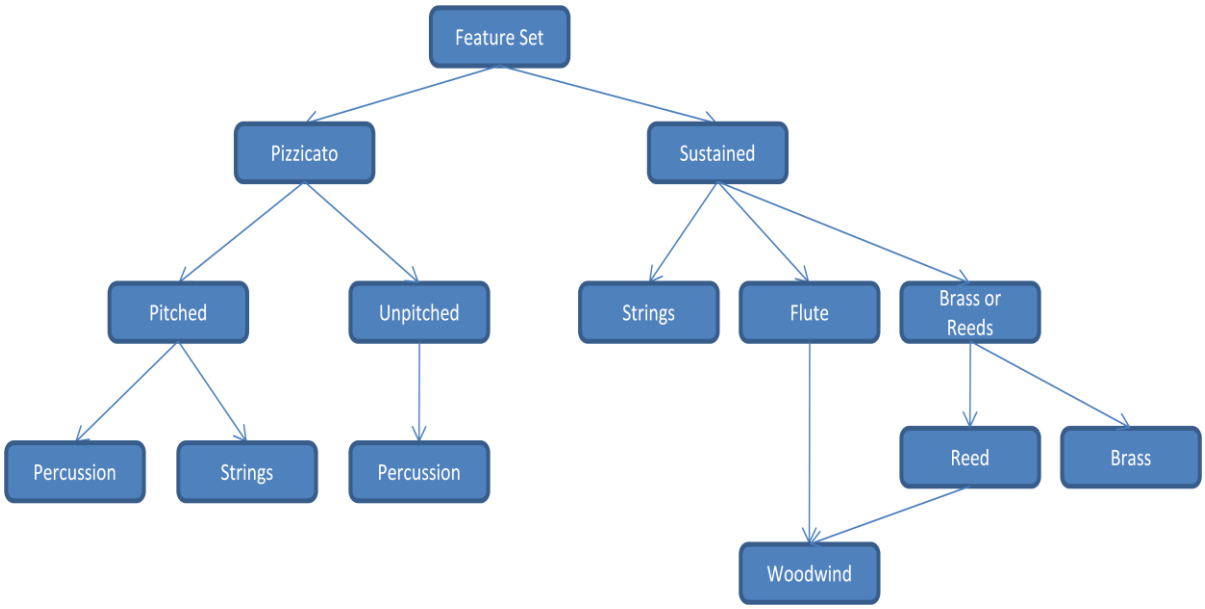

Figure 3 - k-Nearest Neighbor Classification Hierarchy

Some drawbacks from this implementation are the large memory requirements to store the "learned" data set and it also requires a lot of processing power to run through several learned sets and compare against the incoming set.

Implementation

Due to lack of resources this model was designed strictly in software on a windows PC platform. Three modules shown in Figure 4 - Instrument Recognition Diagram comprise the Instrument Recognition AI. Wav sound files were used in place of a live instrument feed. This sound is fed into a feature extraction module to retrieve the spectral and temporal sound components.

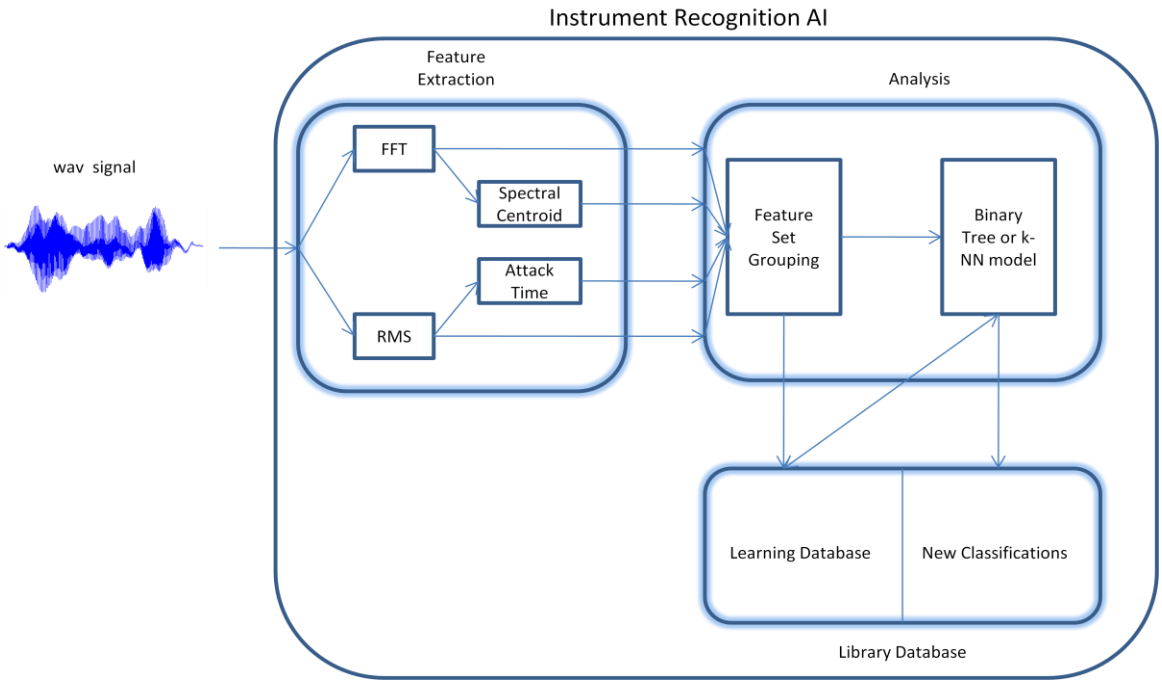

Figure 4 - Instrument Recognition Diagram

Matlab was used for all feature extractions because it has built in functions for performing DFTs and FFTs which helps to reduce the coding complexity. The following figures show the data involved in the features extraction. Figures 5 - 11 show the spectral features and figures 12 - 18 show the temporal feature.

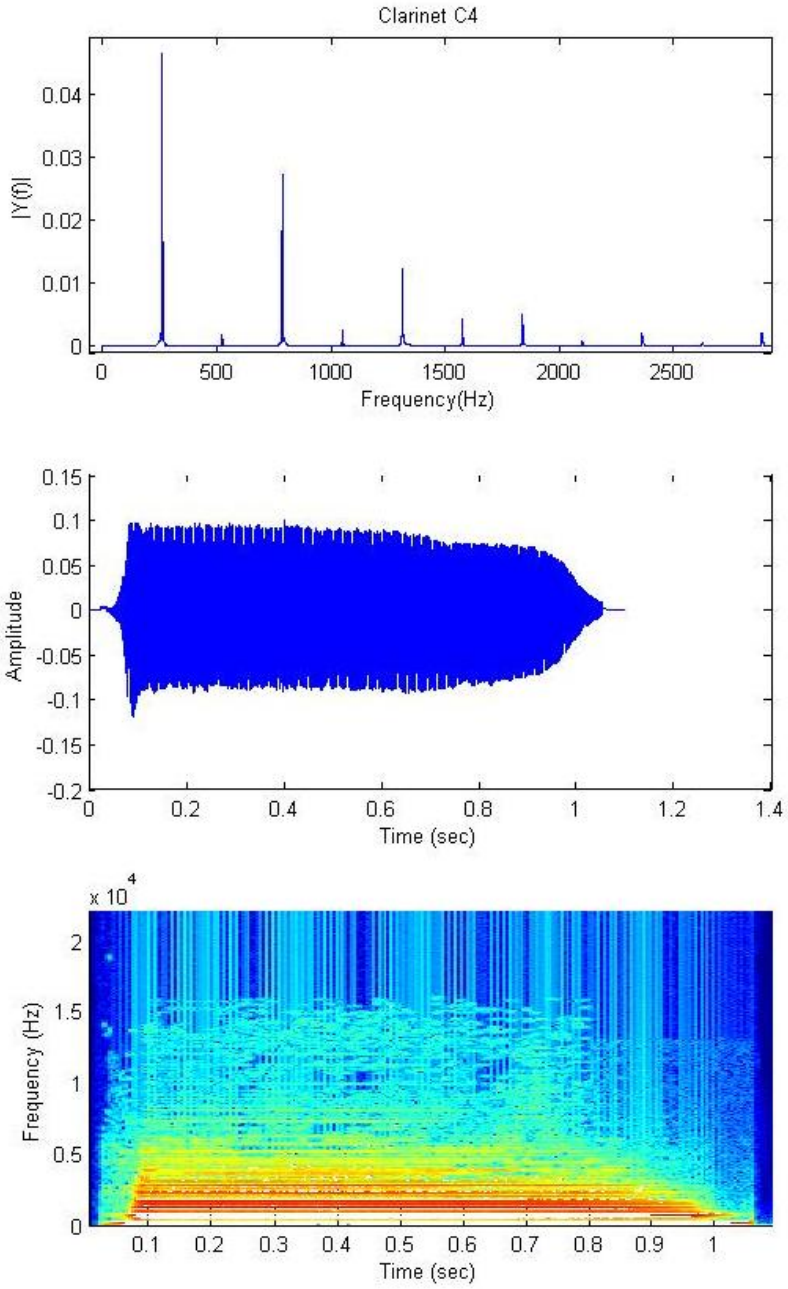

Figure 5 - Spectral Features : Clarinet

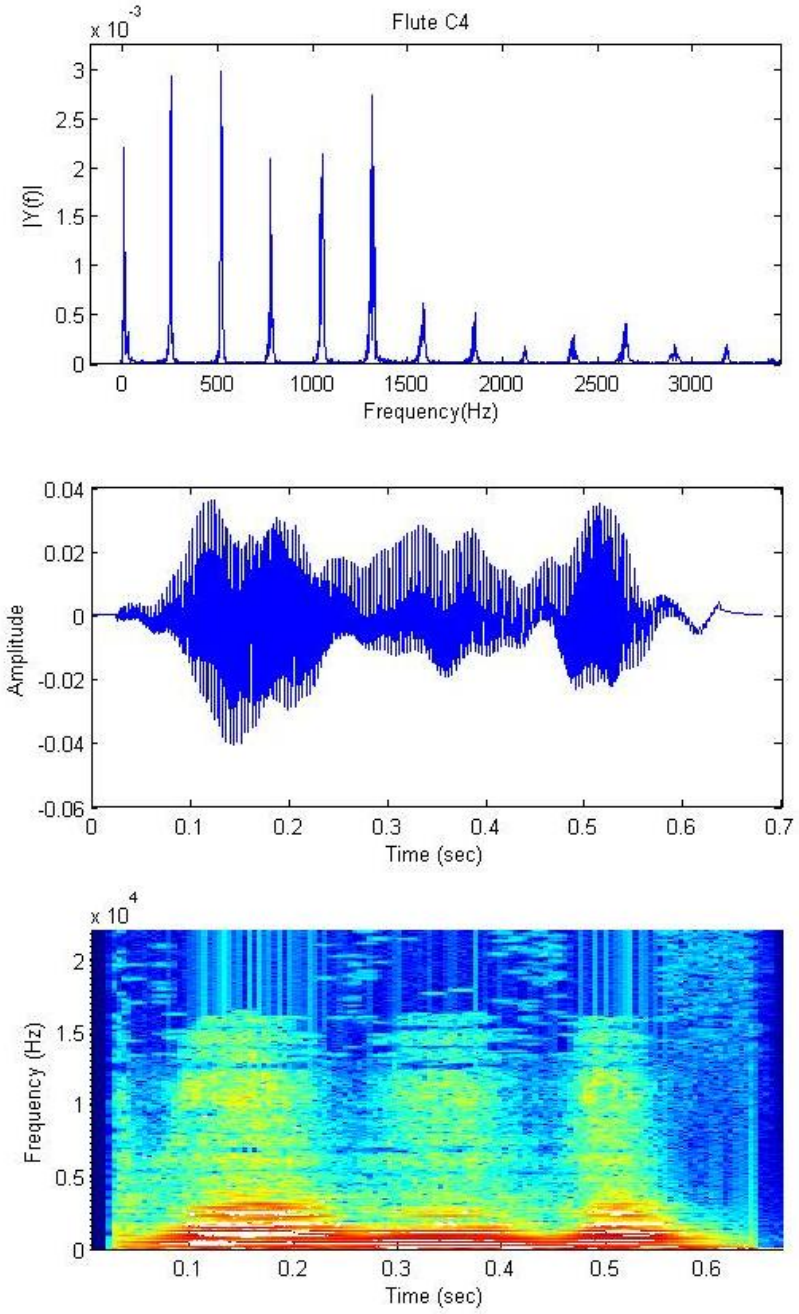

Figure 6 - Spectral Features: Flute

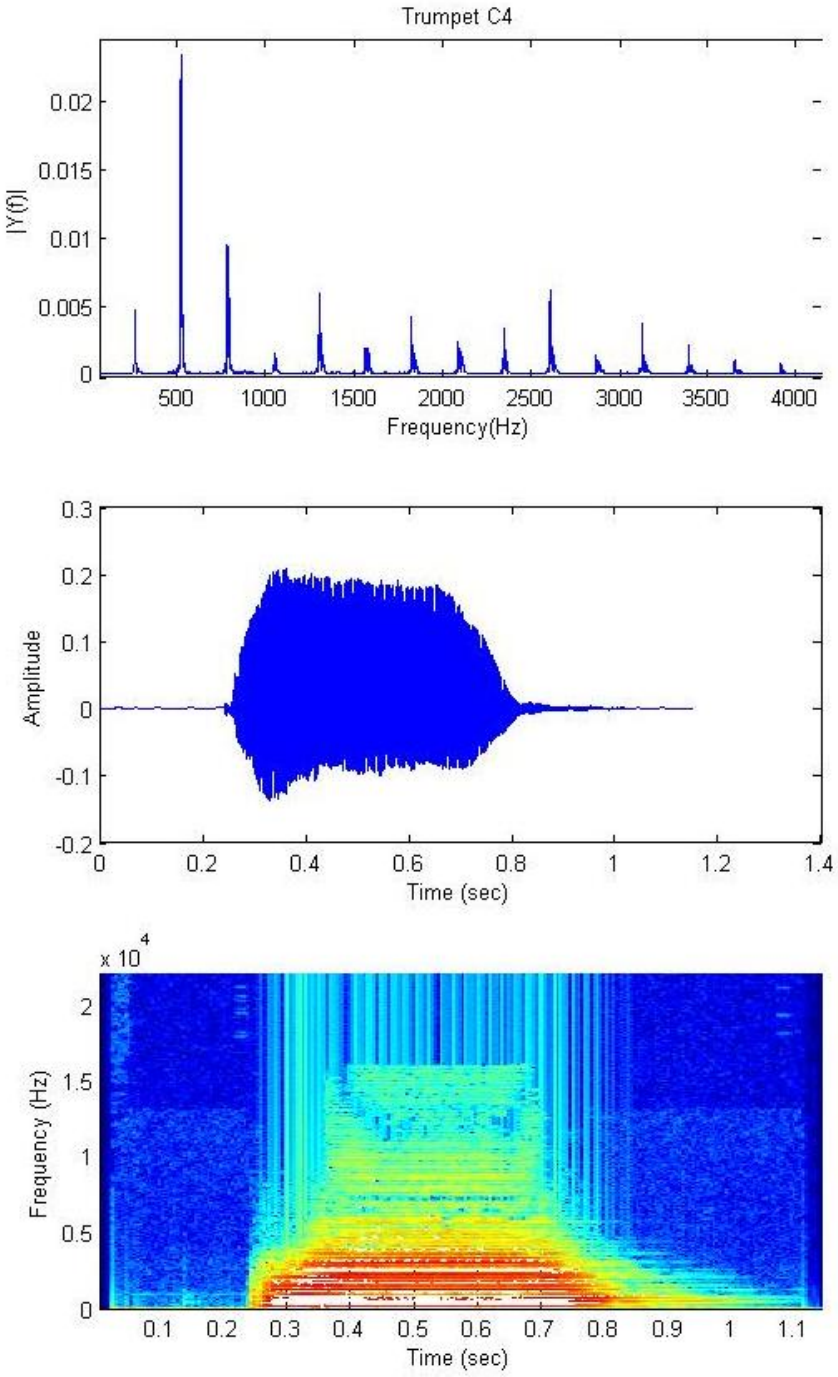

Figure 7 - Spectral Features: Trumpet

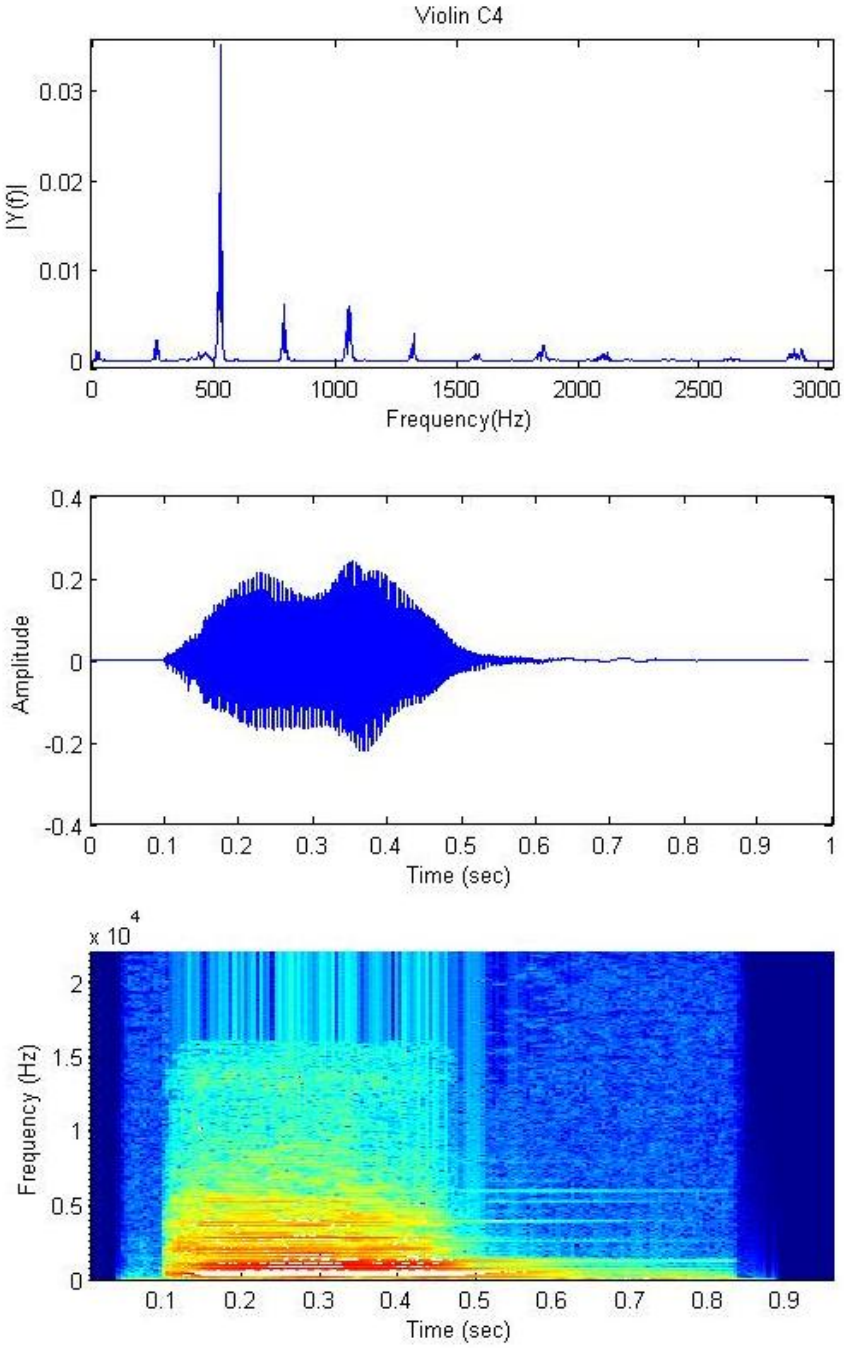

Figure 8 - Spectral Features: Violin

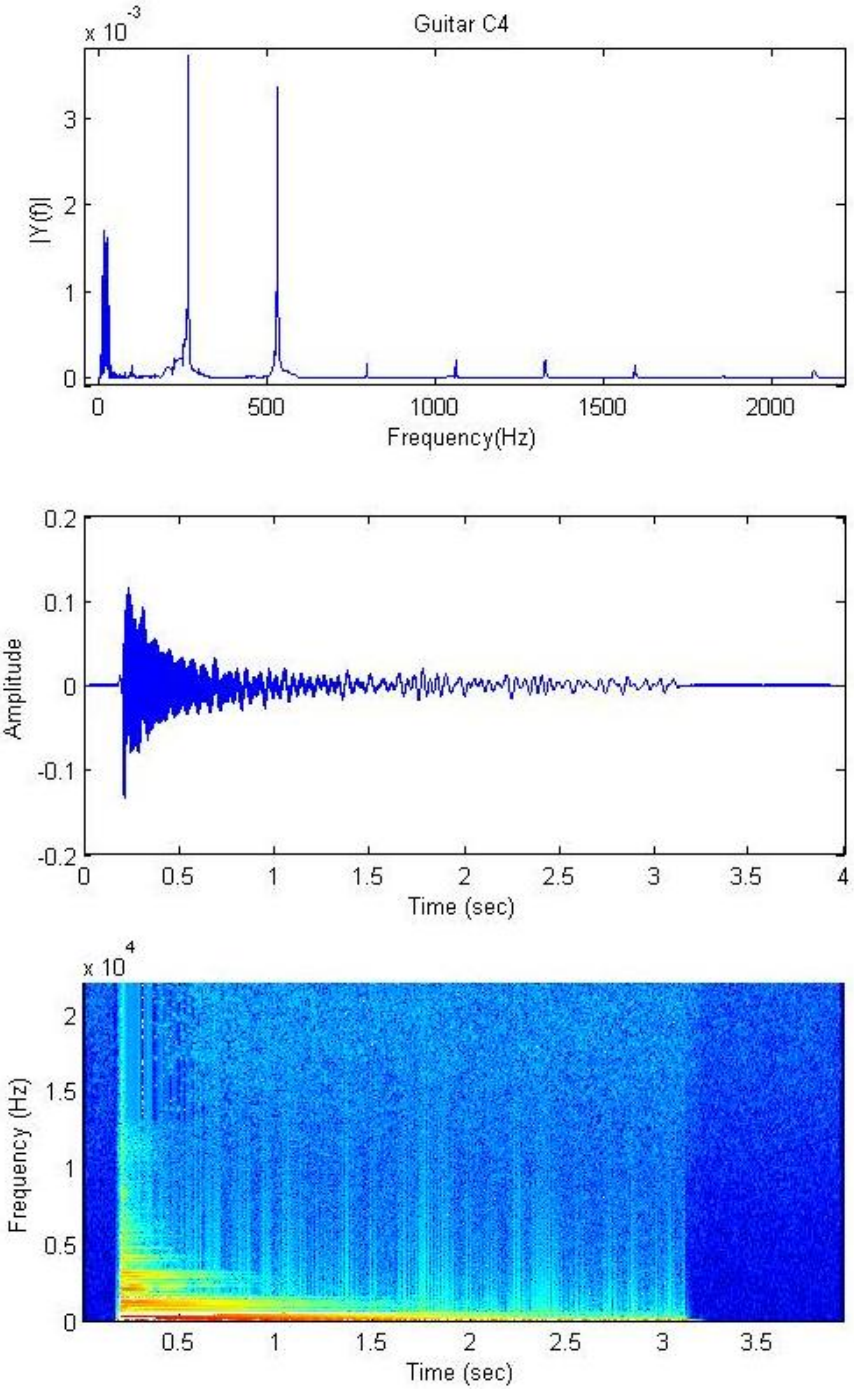

Figure 9 - Spectral Features: Guitar

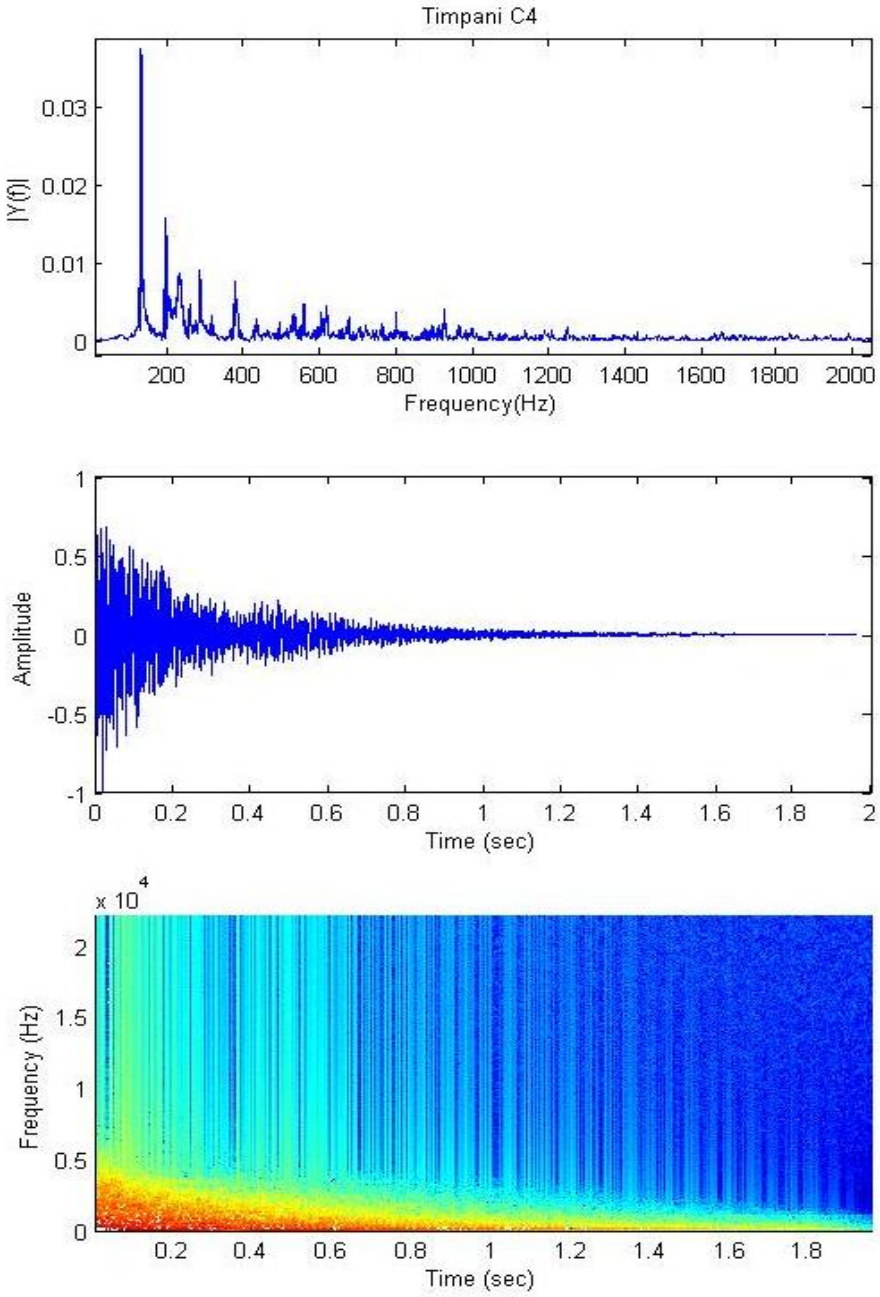

Figure 10 - Spectral Features: Timpani

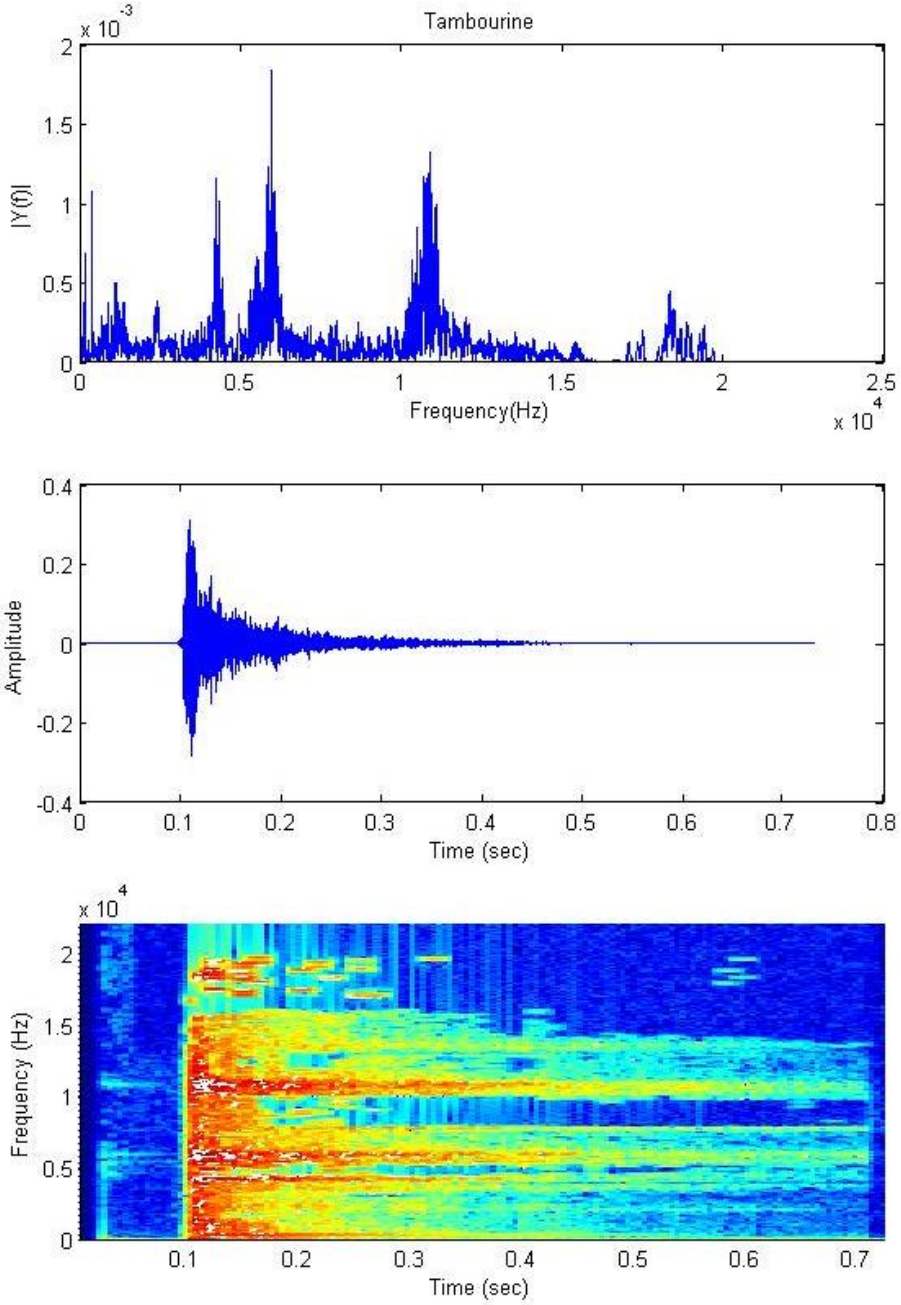

Figure 11 - Spectral Features: Tambourine

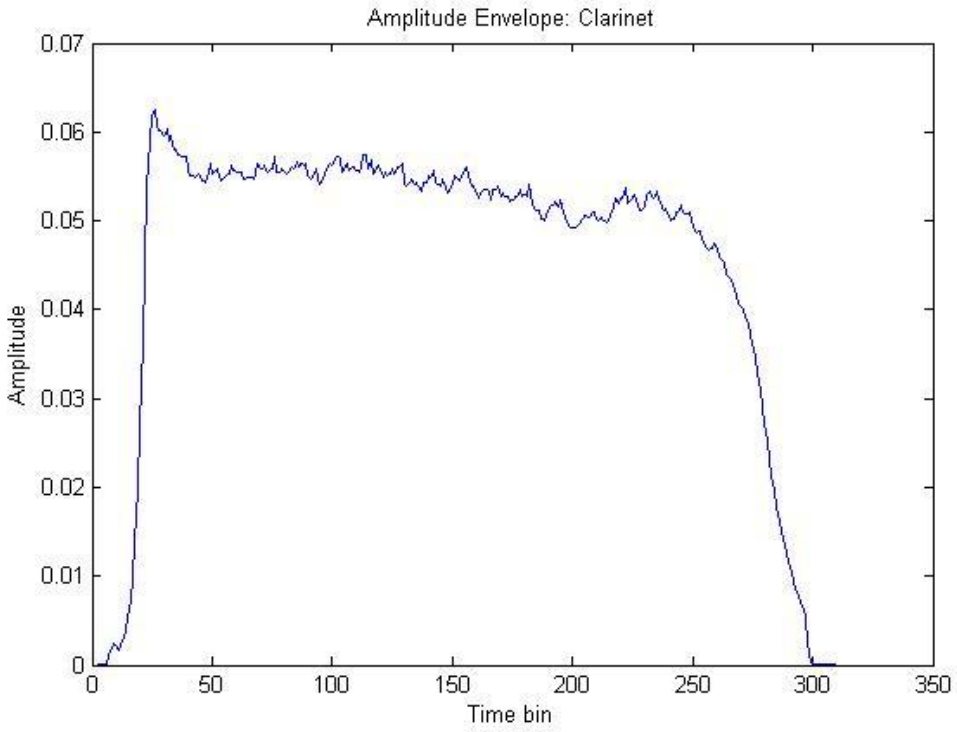

Figure 12 - Temporal Features: Clarinet

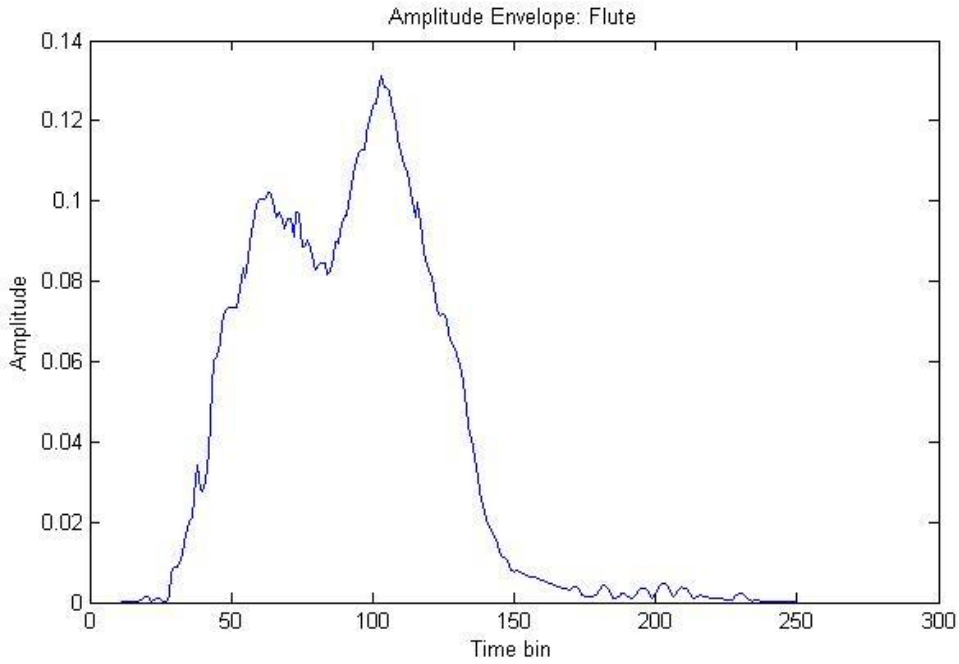

Figure 13 - Temporal Features: Flute

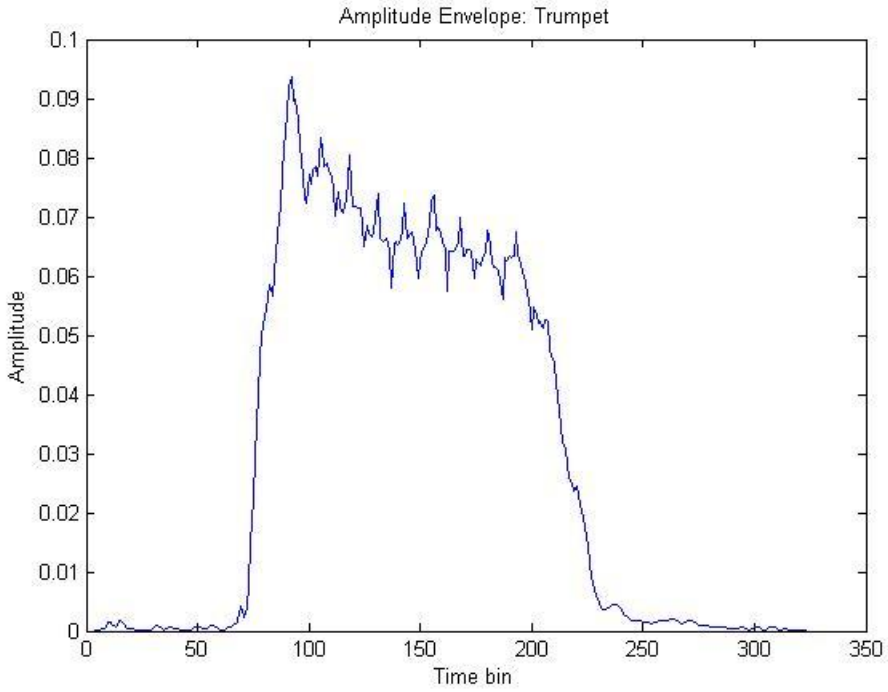

Figure 14 - Temporal Features: Trumpet

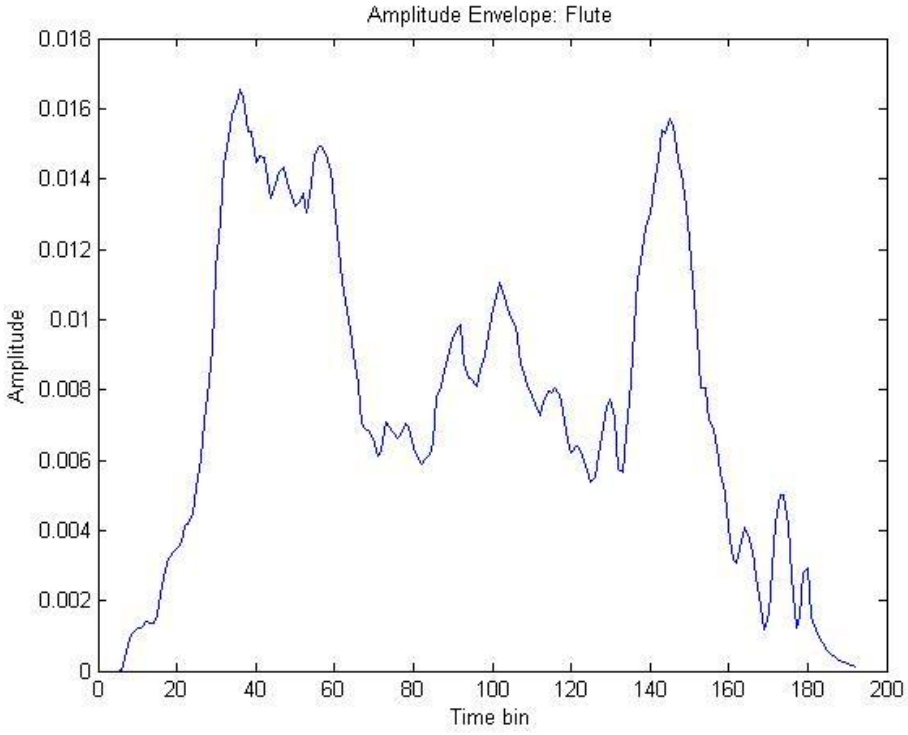

Figure 15 - Temporal Features: Violin

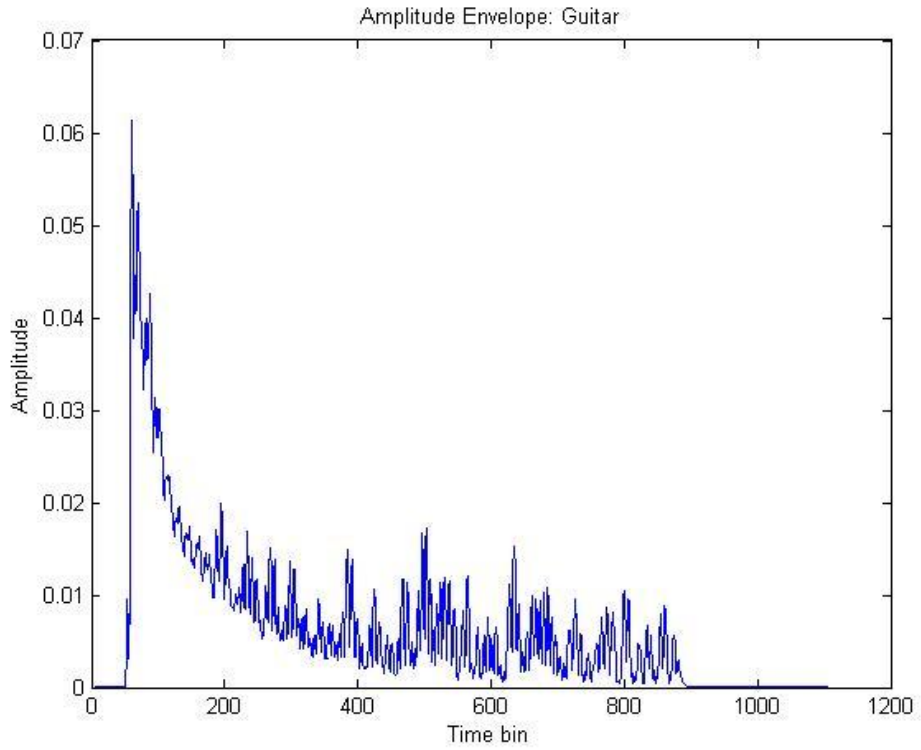

Figure 16 - Temporal Features: Guitar

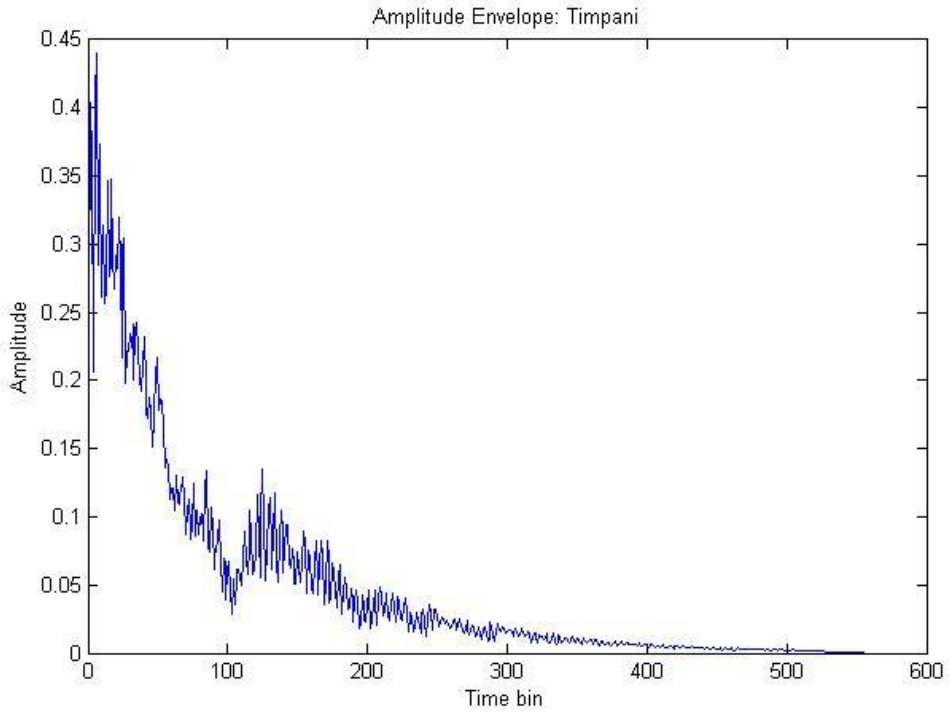

Figure 17 - Temporal Features: Timpani

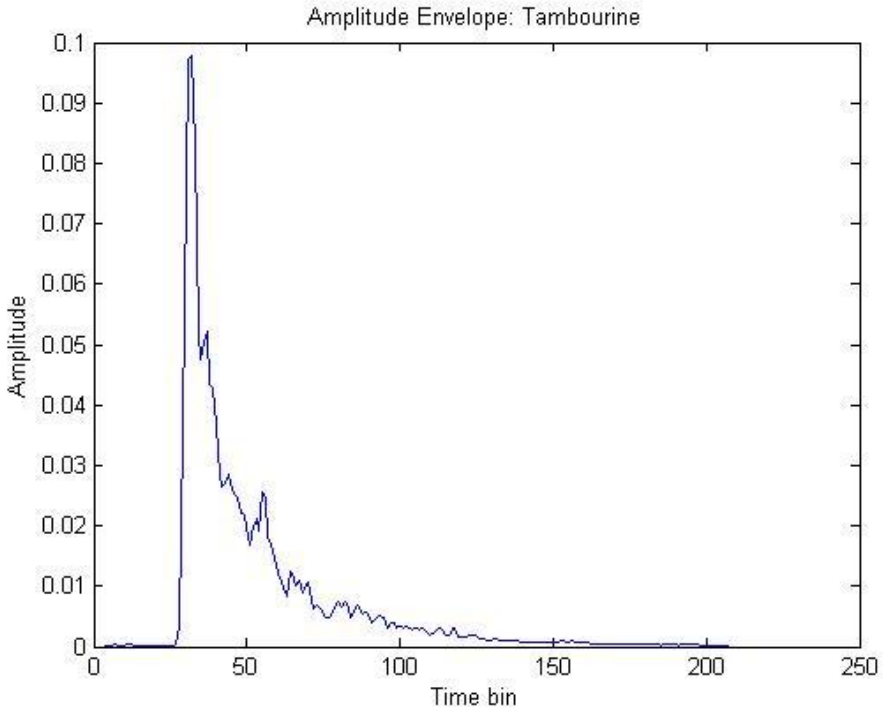

Figure 18 - Temporal Features: Tambourine

The feature extraction uses the data shown in the figures above to extract four features used in our pattern recognition. Attack Time, Harmonics, Spectral Centroid, and the Amplitude Envelope. These features will be used to determine the class of instrument that the sound came from.

For example, string instruments typically have complex resonant properties and very long attack transients; in contrast, brass instruments have a single resonance and much shorter attacks. Woodwinds have short attacks and resonant properties of intermediate complexity.

The next step in implementation is to develop either a binary decision tree or the k-NN model for analyzing the sound features and determining if it is a match to one of the groups it has established during the learning phase in which it creates a database of previous sounds to use for comparison.

Lessons Learned

Originally I thought I had a good start on this project when I had setup a time at the university nearby to collect samples of instruments with varying pitch ranges, loudness, and duration. Unfortunately when it came time, this fell through because of sudden work related travel and schedules weren't able to be realigned to make that happen. After some time I was able to locate some adequate orchestral sound samples online. I didn't have a back up planned at that time which caused some delays getting resources together.

Running solo on this project meant a lot of work to do. Because of scheduling conflicts I had felt best to attempt my own project. Not much I can control here.

One major impact to the project completion was being cut off suddenly from my software resources. I had been using software licenses from my company to complete my project tasks. During one trip I was unable to remote in to use the resources. Upon return I found that all external links had been disconnected. This left three days before my next trip to complete as much of the project as possible before it was due since the company network

was to remain disconnected from outside preventing remote logging. Again this was beyond my control. I could have possibly purchased my own licenses but funds wouldn't support it.

Conclusion

In exploring different areas of sound analysis I've found that certain features of sound are inconclusive in trying to identify the instrument that created it. Features like loudness and pitch aren't unique identifiers. Whereas attack time is a very valuable feature in separating the sound source.
Even though I was unable to complete the project fully it has been a great insight into different methods that could be implemented to achieve a musical instrument recognition program. I would have liked to have delved in further and explored a larger variety of spectral and temporal features and go a step further and create a good discriminatory system that could be used for discerning individual instruments and not just instrument families.

References

[1] T. H. Park, Towards Automatic Musical Instrument Timbre Recognition, Ph.D. Thesis, Princeton University, 2004.

[2] Magneto Technologies [Internet]. 2011 [updated 2011 Apr 24; cited 2011 Apr 23]. Available from: http://www.philharmonia.co.uk/thesoundexchange/make_music/samples/library/

[3] Zhang, Xin and Ras, Zbigniew W. (2007). Analysis of Sound Features for Music Timbre Recognition. *IEEE CS International Conference on Multimedia and Ubiquitous Engineering* (MUE 2007). April 26-28, Seoul, Korea, pp. 3-8.

[4] Zhang, X. and Ras, Zbigniew W. (2006). Sound Isolation by Harmonic Peak Partition For Music Instrument Recognition, *Fundamenta Informaticae Journal Special issue on Tilings and Cellular Automata*, IOS Press, pp. 612-628.

[5] Adam Tindale, Ajay Kapur, and Ichiro Fujinaga. "Towards Timbre Recognition of Percussive Sounds." In Proceedings of the 2004 International Computer Music Conference (ICMC2004), Miami, Fl., 2004.

[6] Alm, Jeremy F. and Walker, James S. (2002) Time-Frequency Analysis of Musical Instruments, *Society for Industrial and Applied Mathematics*, vol. 44, no. 3, pp. 457 – 476.

[7] Slim Essid, Ga el Richard, and Bertrand David. Musical instrument recognition based on class pairwise feature selection. In 5th International Conference on Music Information Retrieval (ISMIR), Barcelona, Spain, October 2004.

[8] Eronen, A., & Klapuri, A. (2000). Musical instrument recognition using cepstral coefficients and temporal features. In: *Proceedings IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP) 2000*.

[9] Fragoulis, D. K., Avaritsiotis, J. N., and Papaodysseus, C. N. Timbre recognition of single notes using an ARTMAP neural network. Proc. of the 6th IEEE International Conference on Electronics, Circuits and Systems. Paphos, Cyprus. 1999.

[10]

T. H. Park, Towards Automatic Musical Instrument Timbre Recognition, Ph.D. Thesis, Princeton University, 2004.

[11]

L. G. Martins, J. J. Burred, G. Tzanetakis, and M. Lagrange. Polyphonic instrument recognition using spectral clustering. In Proc. of ISMIR - International Conference on Music Information Retrieval, 2007.

[12]

Zhang, X. and Ras, Z.W. (2006B). Sound Isolation by Harmonic Peak Partition For Music Instrument Recognition, Fundamenta Informaticae Journal Special Issue on Tilings and Cellular Automata, IOS Press, 2006

[13]

Park T. H. 2000. Salient Feature Extraction of Musical Signals. Master's Thesis. Dartmouth College, Electro-acoustic Music Program.