

Towards Timbre Recognition of Percussive Sounds

Adam Tindale, Ajay Kapur, and Ichiro Fujinaga
Music Technology Department, McGill University
[tindale, ich]@music.mcgill.ca
Electrical and Computer Engineering, University of Victoria
ajay@ece.uvic.ca

Abstract

The development of computer algorithms for music instrument identification and parameter extraction in digital audio signals is an active research field. A musician can listen to music and instantly identify different instruments and the timbres produced by various playing techniques. Creating software to allow computers to do the same is much more challenging. This project will use digital signal processing and machine learning techniques to differentiate snare drum timbres produced by different strike positions and strike techniques.

1 Introduction

There have been many studies into instrument classification (Herrera, Peeters, and Dubnov 2003) and percussion instrument classification (Gouyon and Herrera 2001; FitzGerald, Coyle, and Lawlor 2002; Silpanpää 2000). This study is an attempt at providing additional tools that classify the subtle differences in the timbre that a snare drum can produce. In particular, this paper will investigate differences in snare drum timbres produced by different strike positions and strike techniques.

Instrument classification systems try to classify the signal into a particular timbre space that represents a family of instruments or particular instruments. These systems do not provide provisions for the different timbres produced by a single instrument. However, these systems are able to place a timbre produced by an instrument into a feature subspace. Therefore it must be possible to examine the subspace and learn to identify the different timbres represented within the subspace.

By identifying these different timbres it will be possible to use this information to allow other systems to provide more meaningful information to the user. An interesting possibility is to find ways to use this data in a realtime context. In order to allow for the possibility of realtime recognition of these timbres there must be careful consideration of the fea-

ture extraction algorithms. The eventual goal is to recognize the recognition data as control parameters in order to eliminate the need for sensors to capture snare drum strike position.

2 The Snare Drum

The feature that distinguishes the snare drum from other drums is the snares, which are wires stretched across the bottom head of the drum that vibrate in sympathy when the drum is struck. The snare drum can be struck at different points along the radius of the batter head to produce different timbres because different modes of the membrane are excited and dampened (Rossing 2000). The snare drum may also be struck with a wire brush, which creates a unique timbre, or it can be struck so that the stick makes contact with the snare head and the rim at the same time producing a rimshot.

While there has been some research on snare drums, very little of it deals directly with timbre recognition. The first major published study on the snare drum was mostly concerned with amplitude and durations (Henzie 1960). The study scientifically introduced the idea of a stroke height, the height that the stick starts its strike from, as being the major factor in the resulting amplitude of the strike. Complex interactions between the modes have been observed and discussed (Zhao 1990), which is useful evaluating what type of features to look for when trying to classify timbre. An empirical study (Lewis and Beckford 2000) showed that different types of snare drum heads on the same drum can produce varying timbres. Another study of this nature showed spectra from a snare drum with its snares engaged and not engaged (Wheeler 1989) that demonstrated a large difference in timbre.

See Figure 1 for representations of snare drum signals.

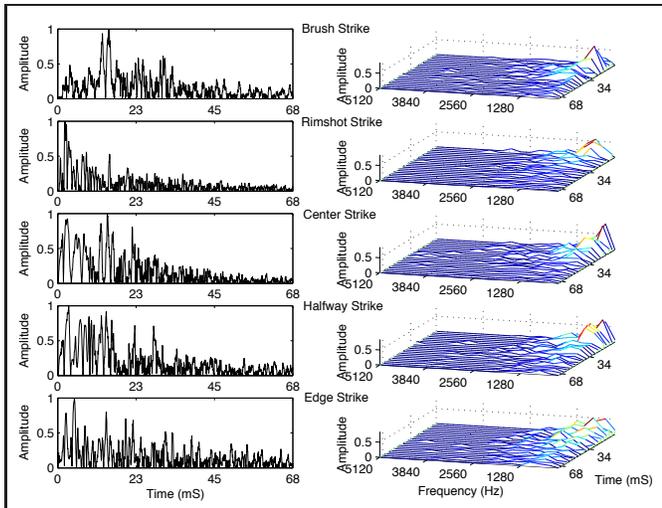


Figure 1: Time Domain and Frequency Domain representations.

3 Software Design

3.1 Feature Extraction Algorithms

Standard features (Herrera, Yetarian, and Gouyon 2002) were examined for their usefulness in classifying subtle differences in timbre of snare drum sounds. One of the main criteria for features is that they are possible to compute in short order so that the system may be later implemented to run in realtime. The features are intended to extract general characteristics of the snare drum in order to be robust to handle different drums and different players.

The features implemented in this study are all calculated in the time-domain, no FFT's are computed in this system. The system uses a square window over the attack section of the sound, which is defined as the point between the onset time and the peak amplitude.

The features were all computed on the attack section of the snare drum sound. The features include: zero-crossings (Gouyon, Pachet, and Delerue 2000), attack time, RMS and temporal centroid. Attack time zero-crossing yields a crude frequency estimation based on the number of zero crossings in the attack section. Attack time yields the actual time (in number of samples) of the attack section. Attack time RMS yields the amount of energy in the attack section. Attack time temporal centroid yields the location of the centre of gravity of the window being examined.

Sub-band analysis techniques were also employed to determine energy in specific bands. Currently, there are four bands: 0–200 Hz, 200–1000 Hz, 1000–3000 Hz, 3000–20,000 Hz. We designed various filter functions to separate each sig-

nal into 4 different subbands. We designed various IIR filters of length 5 to separate the signals into the four different subbands. The energy in each band is measured during the attack phase, which gives a rough estimation as to what modes of the drum are being excited.

3.2 Classification Techniques

Although there are many possible strategies available to classify data, this study uses a feed-forward backpropagation artificial neural network. This classifier was selected because it allows us to use an exemplar-based learning environment in order to draw analogy to the way in which a human learns to label a technique based its resulting timbre. The network is a three layer network with 6 hidden nodes, and 3 output nodes.

4 Results

Snare drum sounds were recorded to a hard drive using a Mark of the Unicorn 896 set to 16-bit resolution and 44.1kHz sampling rate with a Neuman U-87 microphone placed near the edge and suspended perpendicularly over the drum. All strikes were played with an eight-inch stroke height by an expert player. A total of 100 recordings were made: 20 strikes at the center (actually off-center to avoid the "dead spot"), 20 strikes at the edge, and 20 strikes half-way between the edge and the center of the drum, 20 brushes strikes and 20 rimshots. The resulting files were segmented manually and their amplitudes were normalized so that only the timbral qualities were relevant. The files were also run through a gate function which truncated the sample from the previous zero-crossing above $|0.1|$ and then kept only the first 10,000 samples. Leave-one-out cross-validation was used to select testing and training data for evaluating the system.

4.1 Stage 1

Three initial experiments were conducted that were expected to be progressively more difficult. Our first experiment was to distinguish between rimshot, brush strike, and center strike. The second experiment was to distinguish between edge strike, center strike and rimshot. The third experiment was to distinguish between edge strike, halfway strike, and center strike. Subband features were not used in this experiment.

As shown in Table 1, the system was able to successfully classify rimshot and brush strokes with great accuracy. These two timbres are significantly different from standard playing techniques. The standard playing techniques proved to be more difficult to accurately classify but the system was

successful enough that with further investigation into feature selection algorithms it can only increase.

Another experiment was conducted to improve the accuracy of the identification of the standard playing techniques. The subband features were added to the system and experiment 3 was run again. This time center strikes were recognized 90%, middle strikes 95%, and edge strikes 87.5%.

	Exp. 1	Exp. 2	Exp. 3	Exp.4
Center	100%	85%	80%	90%
Halfway	–	–	70%	95%
Edge	–	70%	70%	87.5%
Rimshot	100%	100%	–	–
Brush	100%	–	–	–

Table 1: Results

4.2 Stage 2

Five additional experiments were run with the introduction of some basic spectral domain features. The features used for testing were: spectral centroid, spectral rolloff, spectral flux, mel-frequency cepstrum coefficients, and linear predictive coding coefficients. Some of these features add an FFT component to the system. These features were calculated on a larger window of the sound. As in the previous stage, a gate function was used to determine the onset, which was then used as the beginning point of signal for analysis. The window was ended when the signal dropped below the initial threshold of $[0.1]$.

Table 2 shows the results of several different experiments run with the selections of this feature set. The feature sets for these experiments is outlined below.

1. Spectral Centroid, Spectral Rolloff, RMS
2. RMS, Ramptime, Spectral Centroid
3. Ramptime, Spectral Centroid, Spectral Rolloff, Spectral Flux
4. Linear Predictive Coding Coefficients, RMS, Spectral Centroid
5. Mel-Frequency Cepstrum Coefficients

Although none of these experiments provide outstanding results they use minimal features to achieve their task. Interestingly, the Mel-Frequency Cepstrum Coefficients provide satisfactory results on their own. The lower results of this phase of experimentation suggests that the initial window size is more appropriate for discriminating between the subtle timbres produced by the snare drum.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
Center	85%	90%	90%	60%	80%
Halfway	95%	75%	70%	80%	55%
Edge	85%	85%	75%	85%	70%
Rimshot	85%	90%	90%	65%	85%
Brush	65%	75%	50%	70%	80%

Table 2: Spectral Features Included

5 Future Research

Given the very encouraging results observed in this research, further experiments are planned to investigate deeper into the subtle percussion timbre space. A larger data set of snare drum recordings is being planned that will include multiple players on multiple drums. The problem of recognising the timbres produced at different strike points along the snare head seems to be more complex as the step sizes gets smaller. Along with increasing the recognition rate of the timbres investigated in this study, smaller step sizes will be examined. The study will continue to shrink the step sizes until they become so similar that the system can no longer distinguish between them.

The results of this research will then be applied towards other percussion instruments, beginning with tabla, to begin investigating strategies for timbre recognition systems that can differentiate multiple instruments as well as classify their timbre. As the system improves it is hoped that it can be included into broader recognition systems (Fujinaga and MacMillan 2000; Martin and Kim 1998) thus creating a new type of transcription system that is not only able to label the instrument but also provide timbral information about the musical signals' components.

The performance of different windowing techniques will be evaluated more thoroughly throughout the course of this study. Along with the variable window size demonstrated in this paper, fixed width and multiple windowing techniques will be investigated. More spectral and time domain features will be included and tested for their effectiveness in classification.

6 Conclusion

Our initial experiments into timbre classification of snare drums provides promising results which suggest that with further development and investigation better results can be achieved. As it becomes more apparent which features are major contributors, the system will be modified so that it will be able simply and accurately classify subtle differences in

snare drum timbres. The system provided in this paper offers a classification system which can classify in realtime since the features are calculated in a very small window. Hopefully this research will demonstrate to larger projects that not only is it possible to identify instrument types, but that it is possible to classify the different timbres produced by an instrument.

References

- FitzGerald, D., E. Coyle, and B. Lawlor (2002). Sub-band independent subspace analysis for drum transcription. *Proceedings of Workshop on Digital Audio Effects*, 65–9.
- Fujinaga, I. and K. MacMillan (2000). Realtime recognition of orchestral instruments. *Proceedings of the International Computer Music Conference*, 141–3.
- Gouyon, F. and P. Herrera (2001). Exploration of techniques for automatic labeling of audio drum tracks' instruments. *Proceedings of MOSART: Workshop on Current Directions in Computer Music*.
- Gouyon, F., F. Pachet, and O. Delerue (2000). On the use of zero-crossing rate for an application of classification of percussive sounds. *Proceedings of Workshop on Digital Audio Effects*.
- Henzie, C. (1960). *Amplitude and duration characteristics of snare drum tones*. Ed.D. Dissertation. Ph. D. thesis, Indiana University.
- Herrera, P., G. Peeters, and S. Dubnov (2003). Automatic classification of musical instrument sounds. *Journal of New Music Research* 32(1), 3–21.
- Herrera, P., A. Yetarian, and F. Gouyon (2002). Automatic classification of drum sounds: A comparison of feature selection and classification techniques. *Proceedings of Second International Conference on Music and Artificial Intelligence*, 79–91.
- Lewis, R. and J. Beckford (2000). Measuring tonal characteristics of snare drum batter heads. *Percussive Notes* 38(3), 69–71.
- Martin, K. and Y. Kim (1998). Musical instrument identification: A pattern-recognition approach. *Presented at the 136th meeting of the Acoustical Society of America*.
- Rossing, T. (2000). *The science of percussion instruments*. River Edge, NJ: World Scientific.
- Schloss, A. (1985). *On the Automatic transcription of percussive music: From acoustic signal to high-level analysis*. Ph. D. thesis, CCRMA, Stanford University.
- Silpanpää, J. (2000). Drum stroke recognition. Technical report, Tampere University of Technology.
- Wheeler, D. (1989). Focus on research: Some experiments concerning the effect of snares on the snare drum sound. *Percussive Notes* 27(3), 48–52.
- Zhao, H. (1990). Acoustics of snare drums: An experimental study of the modes of vibration, mode coupling and sound radiation patterns. Master's thesis, Northern Illinois University.