RETRIEVAL OF PERCUSSION GESTURES USING TIMBRE CLASSIFICATION TECHNIQUES

Adam Tindale
Music Technology
McGill University

Ajay Kapur
Electrical Engineering
University of Victoria

George Tzanetakis
Computer Science
University of Victoria

Ichiro Fujinaga
Music Technology
McGill University

ABSTRACT

Musicians are able to recognise the subtle differences in timbre produced by different playing techniques on an instrument, yet there has been little research into achieving this with a computer. This paper will demonstrate an automatic system that can successfully recognise different timbres produced by different performance techniques and classify them using signal processing and classification tools. Success rates over 90% are achieved when classifying snare drum timbres produced by different playing techniques.

1. INTRODUCTION

One major goal of music information retrieval is automatic music transcription. There are two main problems to solve in these systems: instrument recognition and translation into a symbolic musical format (e.g., MIDI). Although the instrument labelling typically returned by such systems is adequate in most cases, the inclusion of different timbre produced by a single instrument would be useful for many different applications and studies.

Currently there are many systems that can successfully classify sounds into instrument groupings but none of them examine the timbre space within these groupings [10]. The goal of this project is to develop a system that can identify the subtle differences in timbre produced by an instrument and classify these differences (see Figure 1). The subtle timbre recognition has the potential to aid in other tasks: drummer recognition, gestural control of music, genre classification of music, etc.

The snare drum was chosen as the instrument for this study because it can create many different subtle timbres produced by controlled and quantifiable performance techniques; although subtle timbres can be produced by other instruments it is often difficult to control their production.

Our previous study presented results with a limited amount of data and only time-domain features [16]. The current study includes a much larger amount of test and training data for the classifier, spectral features, and the exploration of different windowing techniques.

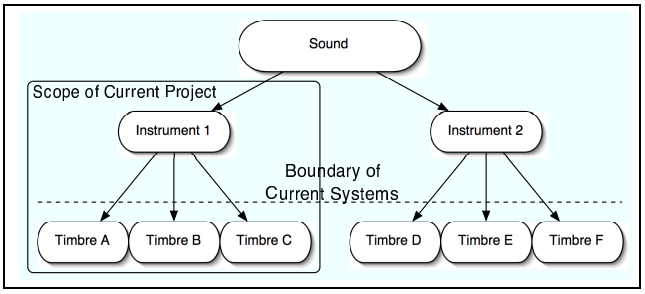

Figure 1. System outline.

2. SNARE DRUM

A snare drum is made up of five main components: the shell, lugs, heads, rims, and the snares. The shell is generally a piece of wood that is bent so that it is a cylinder. Metal fittings, called lugs, are attached to the shell for the purpose of holding all of the components of the drum together. There are two heads on a modern drum: the batter head, that is the top head that is usually struck, and the resonant head, the bottom head. The rims are solid pieces of metal that hold the heads to the drum by applying pressure to the head from the outer edge. The rims can vary the tension on the head so that it may be tuned.The characteristic of the snare drum that differentiates it from other drums is the snares. The snares are usually made of metal and are strung across the bottom head of the drum. The snares vibrate in resonance when the drum is struck adding a noise component to the sound produced by the drum [14].

See Figure 2 for representations of snare drum signals.

3. RELATED RESEARCH

While there has been some research on snare drums, very little of it deals directly with timbre production. The majority of the research falls into two areas: acoustics and automatic transcription. The first major published study on the snare drum was mostly concerned with amplitude and durations [7]. 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 am-

plitude of the strike (see section 5). Complex interactions between the modes of vibration of the heads have been observed and discussed [22], which is useful evaluating what type of features to look for when trying to classify timbre. An empirical study [11] 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 [20] that demonstrated a large difference in timbre. Dahl [1] has done investigations in determining the strike force by analysis of the drumsticks using video capture techniques.

There has been a significant amount of research in automatic music transcription that has dealt specifically with drums. Most of this research deals with elements of the drumset and related percussion instruments in pop music. The earliest major study was conducted by Schloss in his doctoral dissertation where he was able to recognise timbres produced by conga drums and then produce a transcription [15]. Herrera et al. have conducted extensive studies that recognise percussive sounds produced by a drumset instrument [9, 6].

Also studies on extracting timbre information from instruments to be used for real-time control of computers for musical applications exist. The extraction of control features from the timbre space of the clarinet is explored in [3]. Deriving gesture data from acoustic analysis of a guitar performance is explored in [17].

4. IMPLEMENTATION

Two different software packages are employed for this project: Weka [21] and Matlab [13]. Matlab was used to perform the feature extraction and the results are put into a data file for Weka which was used for the classification experiments.

4.1. Feature Extraction

Feature extraction was accomplished with Matlab by creating functions that operate on the audio signals and produce results that are stored as a feature matrix in the Weka file format.

Two main types of features are extracted: time-domain and spectral domain features.

The time-domain features included: Temporal Centroid, Attack Time, RMS, Zero-Crossing Rate [5], Subband Analysis (RMS in four bands:0–200Hz, 200–1000Hz, 1000–3000Hz, 3000–20000Hz).

The spectral domain features included: Spectral Flux, Spectral Rolloff, Spectral Centroid, Spectral Kurtosis, Spectral Skewness, Mel-Frequency Cepstrum Coefficients[12], Linear Predictive Coding Coefficients, Energy in nine wavelet bands, Variance from the mean in each wavelet band [18, 19].

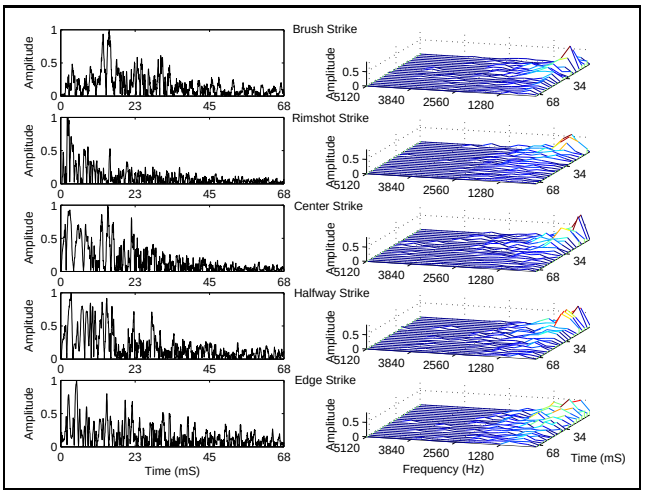

Figure 2. Representations of snare drum sounds.

4.2. Classification

Artificial neural networks are used to classify the data [2]. The network is comprised of six hidden nodes and one output node for each class in the test, which varied as explained below. The net was trained with 1000 epochs.

Different combinations of features and classes are used to train the classifier in order to evaluate the system performance. This study uses seven different types of snare drum strokes in order to create different timbres on the snare drum: rimshot, brush stroke, center, near-center, halfway, near-edge and edge. A rimshot is when the player strikes the rim and head of the drum at the same time (see Figure 3). A brush stroke is when the player hits the drum with a brush instead of a stick. The players are instructed to hit the center of the head when performing both of these strokes. The rest of the strokes are different positions on the batter head. The snare drums are marked so that the players would strike the same place when performing the strokes. Ten-fold cross-validation was used to randomly select samples from the data set as training and testing data [2].

Four different groups of classes are used to train the classifier:

1. All Classes (All)
2. Center, Near-Center, Halfway, Near-Edge, Edge (Only 5) (see Figure 4 for their location)
3. Rimshot, Brush, Edge (RBE)
4. Center, Halfway, Edge (CHE)

5. EXPERIMENT

Three expert players each played three different drums, striking each of the seven different stroke types twenty times. This resulted in 1260 individual samples that are used as testing and training data for the classifier. All players used the same wooden stick (Vic Firth Concert)

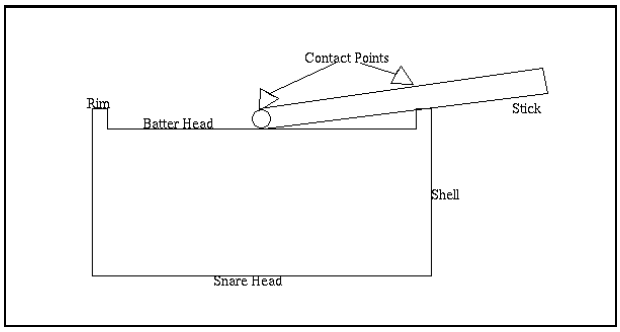

Figure 3. Rimshot

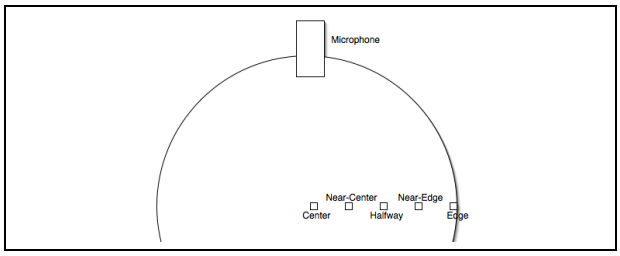

Figure 4. Positions

with a stroke height of six inches [7]. The brush used was a Vic Firth standard brush fully extended. The three snare drums are: Yamaha Concert 14" x 6.5", Ludwig Concert 14" x 6.5", and a Gretsch Standard 14" x 4.5". The drums use standard plastic heads on both the batter and snare head. The heads are manufactured by Remo on the Ludwig and Gretsch drums, and by Yamaha on the Yamaha drum.

Samples were recorded with a Shure SM57 microphone into a Mark of the Unicorn 896 at CD quality (16-bits / 44100Hz). The microphone was placed one inch above the edge of the drum angled down at approximately 30°. The recordings were conducted in the Electronic Music Studio at McGill University that has some sound-proofing.

5.1. Preprocessing

All of the samples were recorded separately and then normalized for input into the system. A gating function was used to determine the position of the onset. The gate function was set to a threshold of -60*dB*. The onset is determined as the first zero-crossing previous to the index returned by the gate function.

The features are calculated on four different window lengths: Attack section (see below), 512, 1024, and 2048 samples from the onset. The "attack" section is defined as the signal from the onset of the sound, as given by the gate function, to point of maximum amplitude. The average length of this window is 610 samples with a standard deviation of 65. The fixed-length windows were calculated beginning from the onset determined by the gate function.

6. RESULTS

All of the combinations of classes and feature sets were collected into separate Weka files and then run through the classifier. High success rates were achieved for all tests. The classifier was able to accurately (greater than 95%) classify the timbres for the tests with three classes (RBE & CHE). The tests with larger number of classes also performed with great accuracy (see Table 1 and 2). Many of the misclassifications in these tests were classified as the next nearest timbre.

By using Weka, k-nearest neighbour (kNN) (see Table 3 and 4) and Support Vector Machines (SVM) (see Table 5 and 6) classifiers were examined.

Timbres	All	Only 5	RBE	CHE
Attack	**89.3%**	**88.3%**	**99.8%**	98.1%
512	85.0%	86.1%	99.4%	96.4%
1024	85.6%	86.6%	99.4%	98.3%
2048	85.3%	85.6%	99.4%	**98.9%**

Table 1. Results using all features with different number of classes and different window lengths using Neural Network classifier.

Timbres	All	Only 5	RBE	CHE
Attack	73.1%	71.1%	96.8%	91.3%
512	**79.9%**	77.0%	98.1%	95.9%
1024	77.1%	**79.2%**	98.1%	97.2%
2048	79.1%	79.0%	**99.3%**	**98.1%**

Table 2. Results using only the time-domain features with different number of classes and different window lengths using Neural Network classifier.

Timbres	All	Only 5	RBE	CHE
Attack	**94.9%**	**95.3%**	98.1%	**99.3%**
512	93.0%	89.6%	98.9%	96.1%
1024	94.4%	94.1%	**99.4%**	98.7%
2048	92.6%	91.1%	99.3%	96.9%

Table 3. Results using all features with different number of classes and different window lengths using kNN classifier.

The kNN classifier was the most consistent classifier, nearly all of its results are above 90%. The SVM classifier performed very well with the three classes but poorly with the larger sets of classes using time-domain features. The neural network had the highest result (99.8% with all features on the RBE class set).

The different sets of features yielded interesting results. The time-domain features performed nearly as well as the full feature set when classifying the groups with only three classes but significantly less effective when classifying

Timbres	All	Only 5	RBE	CHE
Attack	90.8%	**90.9%**	96.1%	95.7%
512	90.9%	88.8%	98.9%	**97.2%**
1024	91.2%	87.6%	**99.1%**	96.9%
2048	**92.0%**	90.0%	98.9%	97.2%

Table 4. Results using only the time-domain features with different number of classes and different window lengths using kNN classifier.

Timbres	All	Only 5	RBE	CHE
Attack	**86.6%**	**86.8%**	99.3%	97.4%
512	83.4%	79.7%	98.7%	92.6%
1024	82.1%	82.3%	98.1%	**97.4%**
2048	83.5%	82.6%	**99.6%**	96.7%

Table 5. Results using all features with different number of classes and different window lengths using SVM classifier.

Timbres	All	Only 5	RBE	CHE
Attack	57.1%	55.1%	94.3%	85.4%
512	65.4%	59.0%	97.0%	89.1%
1024	68.1%	61.4%	98.0%	91.1%
2048	**68.8%**	**62.1%**	**98.1%**	**91.9%**

Table 6. Results using only the time-domain features with different number of classes and different window lengths using SVM classifier.

multiple classes. These observations suggest that it may be possible to implement an accurate recognition system with only time-domain features, which would allow short-time operations for real-time recognition.

The spectral features were very useful for differentiating the different positions along the radius of the drum. Overall, the classifiers performed the "Only 5" and the "CHE" tests an average of 7.8% better with the spectral features than with the only the time-domain features.

The different window sizes were not major contributers to the overall recognition rate. Further investigation into appropriate windowing techniques will be performed in order to maximise the results. It was interesting that the smallest window (512 samples) was still able to classify as accurately as the long window and that the fitted window ("Attack" window) seemed to have little benefit.

7. CONCLUSION

We have presented a system that can successfully recognise the subtle differences in timbre produced when a snare drum is struck in different locations along the batter head as well as other performance techniques. The present research into the identification of subtle timbres produced by an instrument is a first step towards a comprehensive system that can transcribe music and provide information at the timbral level. Future research will involve applying this system to other instruments in different contexts and then integrating it into an automatic transcription system as well as investigating other features and classification techniques. The system will also be implemented for a real-time system and reevaluated.

Data collected in this study will be made available to interested parties upon contacting the authors.

8. ACKNOWLEDGEMENTS

Thanks to Manj Benning for invaluable help with the Matlab coding. Thanks to D'Arcy Phillip Gray, Kristie Ibrahim and Sarah Mullins for taking the time to play the snare drums for this experiment.

9. REFERENCES

[1] Dahl, S. 2001. *Arm motion and striking force in drumming.* International Symposium on Musical Acoustics. 1: 293–6.

[2] Duda, R., P. Hart, and D. Stork. 2000. *Pattern classification.* New York: John Wiley & Sons.

[3] Egozy, E. 1995. *Deriving musical control features from a real-time timbre analysis of the clarinet.* Masters Thesis, Department of Electrical Engineering and Computer Science. Massachussetts Institute of Technology.

[4] Fujinaga, I., and K. MacMillan. 2000. Realtime recognition of orchestral instruments. *Proceedings of the International Computer Music Conference.* 141–3.

[5] Gouyon, F., F. Pachet, and O. Delerue. 2000. On the use of zero-crossing rate for an application of classification of percussive sounds. *Proceeding of the Workshop on Digital Audio Effects.*

[6] 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.*

[7] Henzie, C. 1960. *Amplitude and duration characteristics of snare drum tones.* Ed.D. Dissertation. Indiana University.

[8] Herrera P., X. Amatriain, E. Batlle, and X. Serra. 2000. Towards instrument segmentation for music content description: A critical review of instrument classification techniques. *International Symposium on Music Information Retrieval.*

[9] Herrera, P., A. Yeterian, 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.

[10] Herrera, P., G. Peeters, and S. Dubonov. 2003. Automatic classification of musical instrument sounds. *Journal of New Music Research* 32 (1): 3–21.

[11] Lewis, R., and J. Beckford. 2000. Measuring tonal characteristics of snare drum batter heads. *Percussive Notes* 38 (3): 69–71.

[12] Logan, B. 2000. Mel-frequency cepstral coefficients for music modeling. *Proceedings of International Symposium on Music Information Retrieval.*

[13] *MATLAB reference guide.* 1992. Natick, MA: The Mathworks, Inc.

[14] Rossing, T. 2002. *The science of sound.* San Francisco: Addison-Wesley Publications Company.

[15] Schloss, A. 1985. *On the Automatic transcription of percussive music - From acoustic signal to high-level analysis.* Ph.D. Dissertation. CCRMA, Stanford University.

[16] Tindale, A., A. Kapur, and I. Fujinaga. 2004. Towards timbre recognition of percussive sounds. *Submitted to ICMC 2004.*

[17] Traube, C., P. Depalle, and M. Wanderley. 2003. Indirect acquisition of instrumental gesture based on signal, physical and perceptual information. *Conference on New Interfaces for Musical Expression.*

[18] Tzanetakis, G., G. Essl, and P. Cook. 2001. Audio analysis using the discrete wavelet transform. *Proceedings of Conference in Music Theory Applications.*

[19] Tzanetakis, G., and P. Cook. 2002. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing* 10 (5): 293–302.

[20] Wheeler, D. 1989. Focus on research: Some experiments concerning the effect of snares on the snare drum sound. *Percussive Notes* 27 (4): 48–52.

[21] Witten, I., E. Frank, and M. Kaufmann. 2000. *Data mining: Practical machine learning tools with java implementations.* San Fransico: Addison-Wesley Publications Company.

[22] Zhao, H. 1990. *Acoustics of snare drums: An experimental study of the modes of vibration, mode coupling and sound radiation patterns.* M.S. Thesis. Northern Illinois University.