SNARE DRUM MOTION CAPTURE DATASET

Robert Van Rooyen
rvanrooy@uvic.ca

Andrew Schloss
aschloss@uvic.ca

George Tzanetakis
gtzan@cs.uvic.ca

University of Victoria, 3800 Finnerty Road, Victoria, BC V8P 5C2

ABSTRACT
Comparative studies require a baseline reference and a documented process to capture new subject data. This paper combined with its principal reference [1] presents a definitive dataset in the context of snare drum performances along with a procedure for data acquisition, and a methodology for quantitative analysis. The multi-volume dataset contains video, audio, and discrete two dimensional motion data for forty standardized percussive rudiments.

Keywords
Dataset, drum, motion capture, rudiments, robotics

1. INTRODUCTION
The production of a quality snare drum performance can take many years of instruction and practice to achieve. Subtle nuances in timing, dynamics, and timbre can separate musicians despite using the same instrument, striking implements, and sheet music. How is this possible? It is a well-known fact that each artist develops their own style, but can this property be quantified in a tangible and reproducible fashion? By recording a performance in a consistent and non-invasive manner, we can in fact discover what makes a particular performer's work unique among their peers.

In order to establish a reference recording, a standardized and well documented score needs to be created or identified. Further, a repeatable data acquisition process is required to create not only the baseline, but subsequent recordings for the comparison set. An analysis of the recordings can lead to new discoveries in human motion as well as highlighting the subtle differences between musicians.

2. MOTIVATION
Studying the complexities of human percussive performance can lead to a deeper understanding of how musicians interpret a musical score while simultaneously imparting personal expressiveness. This knowledge can serve to not only educate other musicians on mastering technique, but also to quantifiably describe what an exceptional performance looks like from a multi-dimensional scientific perspective. Moreover, scalable motion models and machine learning techniques can be developed to render more expressive performances in other mediums, such as robotic instruments. Although this research is being conducted in the context of music, it is conceivable that other branches of study may find elements of the dataset applicable, such as animation or cognitive sciences.

3. PRIOR WORK
Capturing human motion during a performance represents a significant challenge with respect to recording quality data without encumbering the musician or instrument. The research conducted by Tindale, A. et al. [2] established an inventory of techniques and sensors to acquire a wide variety of performance measurements, however many methods proved to alter the sound and/or playability of the instrument, which can negatively impact the quality of the data.

A detailed analysis of percussionist motion was undertaken by S. Dahl, et al. [3] in which the focus was to understand the force and sound level associated with an onset event. The research team used non-invasive video motion capture, but also outfitted each striking implement with LED markers, strain gauges, and conduction paths. The results were highly informative and represent seminal work in the field.

A comprehensive multi-dimensional percussive dataset created by Gillet, O. and Richard, G. [4] known as the "ENST-Drums" was released in 2006. This dataset offers a rich set of audio/video data spanning three professional drummers. All of the data was collected non-invasively and manually annotated with respect to onset time and instrument type. The primary difference in comparison to our dataset is that it provides a macro view of an entire drum kit. Further, the research team used two normal speed (25 frames per second) video cameras as opposed to a high-speed camera on a single instrument with distance calibration.

In contrast to previous studies [2, 3, 4], our work centered on a consistent non-invasive method for acquiring calibrated performance data using a pragmatic approach with commodity off-the-shelf equipment. Further, a concentration on implement tip motion, in the context of complex rudiments, enables a deeper study of statistical distributions across multiple attributes such as timing, velocity, and dynamics.

4. DRUM RUDIMENTS
As an internationally recognized standard compiled by the Percussive Arts Society [5], the 40 rudiments in Table 1 represents a "vocabulary for contemporary percussionists [5]." With a stated goal of capturing a standardized set of recordings that offer rich nuanced performance opportunities, the rudiments provided a well-documented and natural choice. In addition to the rudiments, calibration sequences were recorded for striking implement position calibration and dynamic sound level, which are also included in the dataset.

Table 1. Percussive Arts Society rudiments

1	Single Stroke Roll	21	Flam Accent
2	Single Stroke Four	22	Flam Tap
3	Single Stroke Seven	23	Flamacue
4	Multiple Bounds Roll	24	Flam Paradiddle
5	Triple Stroke Roll	25	Single Flammed Mill
6	Double Stroke Open Roll	26	Flam Paradiddle-diddle
7	Five Stroke Roll	27	Pataflafla
8	Six Stroke Roll	28	Swiss Army Triplet
9	Seven Stroke Roll	29	Inverted Flam Tap
10	Nine Stroke Roll	30	Flam Drag
11	Ten Stroke Roll	31	Drag
12	Eleven Stroke Roll	32	Single Drag Tap
13	Thirteen Stroke Roll	33	Double Drag Tap
14	Fifteen Stroke Roll	34	Lesson 25
15	Seventeen Stroke Roll	35	Single Dragadiddle
16	Single Paradiddle	36	Drag Paradiddle #1
17	Double Paradiddle	37	Drag Paradiddle #2
18	Triple Paradiddle	38	Single Ratamacue
19	Single Paradiddle-diddle	39	Double Ratamacue
20	Flam	40	Triple Ratamacue

5. DATA ACQUSITION AND PROCESSING

Our earlier work [1] described the methodology associated with capturing and processing multi-dimensional performance data, which is composed of video, audio, transducer data, and annotated temporal position information. This process was used for each rudiment in Table 1, which resulted in a large collection of files. Each file in the data set can be reconciled by file name. The naming convention is of the form: “<device> <rudiment> <bpm> [left | right].{mp4, wav, csv}.” The videos are in MPEG4 format with a resolution of 848x480 at 240 frames per second. The audio and transducer tracks were recorded at 16-bit/44.1Khz. Finally, the annotated temporal position information was archived into comma separated value (CSV) files for simple import into MATLAB, Excel, and many other environments given its ubiquity.

In reference to the first rudiment list in Table 1, a “Single Stroke Roll” frame capture of the video, shown in Figure 1, depicts the first strike of the left (red tip) striking implement.

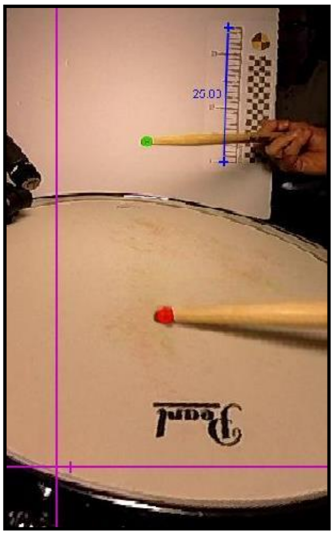

Figure 1. Motion capture frame

A partial plot of the left and right annotations for the first rudiment appears in Figure 2. An offset representing the minimum of the left (green) striking implement has been subtracted from both signals in order to show the relative positions, which due to an intentional drum head angle, results in a z-axis projection. This is illustrated by the right (red) striking implement, whose minimum is approximately 10cm above the left minimum.

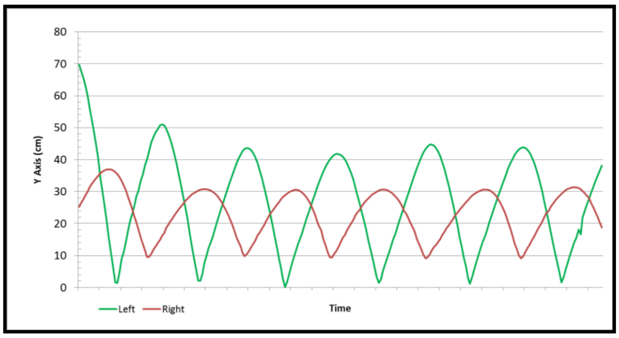

Figure 2. Left and right annotation plot

6. CONCLUSION AND FUTURE WORK

A methodology and real world dataset has been established for current and future research in the field of human motion. The rudiment dataset will be released in several volumes and hosted in the Computer Science Department at the University of Victoria MISTIC lab. In addition, the dataset will be made available for mirroring to enable long-term accessibility.

Our specific research interest is in the creation of musical robots, which also includes the areas of digital signal processing and machine learning. By creating models of human motion, electromechanical devices can begin to move in a more fluid and relatable manner, which is the ultimate goal of this research.

With respect to future work, the rudiment dataset has the potential for multiple applications that include automated transcription, pedagogical studies, digital signal processing, machine learning, and robotics. Therefore, we encourage current and future researchers to leverage and potentially extend this dataset by including additional recordings in order to improve our collective understanding of human musical performances.

7. ACKNOWLEDGEMENT

In addition to multiple conceptual and reference contributions to this paper, Professor Andrew Schloss performed all of the rudiments in the accompanying dataset. As a world renowned percussionist, his experience and attention to detail has provided a high-quality set of definitive recordings for research, which was greatly appreciated by the research team.

8. REFERENCES

[1] R. Van Rooyen and G. Tzanetakis, "Pragmatic Drum Motion Capture System," in *New Interfaces for Musical Expression*, Louisiana, 2015.

[2] A. R. Tindale, A. Kapur, G. Tzanetakis, P. Driessen and A. Schloss, "A Comparison of Sensor Strategies for Capturing Percussive Gestures," in *International Conference of New Interfaces for Musical Expression*, Vancouver, 2005.

[3] S. Dahl, M. Grossbach and E. Altenmuller, "Effect of Dynamic Level in Drumming: Measurements of Striking Velocity, Force, and Sound Level," in *International Conference of the Forum Acusticum*, Denmark, 2011.

[4] O. Gillet and G. Richard, "ENST-Drums: an extensive audio-visual database for drum signals processing," GET / ENST, CNRS LTC1, Paris, 2006.

[5] "Percusive Arts Society," [Online]. Available: http://www.pas.org/index.aspx. [Accessed 27 December 2014].