Snare Drum Performance Motion Analysis

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

Human motion associated with a percussion performance is incredibly complex and filled with subtle nuances that span timing, dynamics, and timbre. Recreating such a performance robotically or through synthesis requires a systematic method of data acquisition and analysis that leads to a concise set of parameters to drive multi-dimensional models. By using statistical analysis to derive parameter vectors that configure stochastic models, we can begin to infuse human qualities in a rendering while simultaneously creating uniqueness in each and every performance. This can translate to a more enjoyable performance for the listener as well as a richer pallet for composers.

Author Keywords

Percussion, motion analysis, rudiments, robotics, stochastic model

ACM Classification

[Mathematics of Computing] Stochastic Processes, [Applied Computing] Sound and Music Computing, [Computing Methodologies] Motion Capture.

1. INTRODUCTION

A percussion performance is composed of striking events over time that includes timing variations (relative to the beat), position changes on the drum head, and changes in striking velocity that proportionally translate to sound pressure level. Taken together as a set, all of the aforementioned variations are unique with each performance given the same musician and score. Such differences can be even more profound across a set of musicians with comparable competence performing the same piece. The former is primarily due to stochastic processes throughout the human body and brain, whereas the latter can be attributed to the addition of subtle stylistic attributes in micro-timing [1].

The goal of this paper is to explore a single representative performance as a case study in the context of timing, velocity, and position in order to identify stochastic and intentional micro timing components. Further, by applying a statistical analysis to onset, velocity, and position, a parameter vector can be derived and used to render a unique performance of a score utilizing a stochastic model that can be coupled with an equally capable robotic or synthesis system. To be clear, the objective is parameter identification, model definition, and a basic understanding of range as opposed to optimization or generalization across a large population of musicians and/or performances.

NIME'16, July 11-15, 2016, Griffith University, Brisbane, Australia.

The goal of this research is to understand and infuse a human touch into robotic percussion performances. Although a robotic rendering can be technically accurate with respect to the piece being played, it often lacks emotion and subtle variety that comes with a multi-dimensional human performance. Despite the fact that robotic musicians may never approach the richness and spontaneity of a human musician, it is possible to direct mechatronic systems to render performances that are more life-like and thus more pleasing to the listener.

Sam Phillips, who arguably invented Rock 'n' Roll, embraced the idea of "perfect imperfection" [2]. As the creator of Sun Records in Memphis Tennessee, he realized that subtle flaws in a performance gave songs a soul. Artists such as Elvis Presley and Johnny Cash capitalized on this approach in countless recordings that proved beneficial to their success and the nearly universal enjoyment of their music. All of this speaks to the power of subtle imperfections that bring life to a recording or live performance.

A fine example of rendering a human like performance on an actual instrument is the Steinway & Sons Spirio high resolution player piano[1]. This system has been designed with the capability of recording all of the subtle keyboard and pedal work that take place in a live performance. Once a given performance has been recorded using their proprietary system, it can be played again and again without losing its virtuosity or most subtle emotion. Reproducing this capability with a percussion instrument represents a unique challenge as the physical constraints and instrument response are quite different. Going beyond playback and introducing stochastic models is yet another level of sophistication that can open new avenues of musical expression. We believe that it is important to capture the details of human percussive performance by analyzing the actual motion patterns used to create the sound rather than solely the sound itself. This requires a sophisticated motion acquisition system which has been described in previous work [3].

2. RELATED WORK

A large body of work exists in the analysis of expressive music performances on a variety of instruments [4, 5, 6, 7, 8, 9]. Research conducted by Berndt and Hehnal explored the degrees of freedom with respect to timing over several feature classes that included human imprecision [10]. Formal timing models were designed and implemented within a MIDI environment, however other attributes such as dynamics and timbre were considered as future work. With regard to randomness, the team cited psychological and motor conditions as the primary contributors to timing accuracy that followed a Gaussian distribution. This was further broken down into macro and micro timing components, with the former being long term tempo drifts and the latter onset delays. The magnitude of timing deviations were subject to musical context and technical difficulty, and quantified using a normal distribution with a

[1] A press release on the Steinway & Sons Spirio player piano can be viewed at http://tinyurl.com/lcbwjxn.

mean of the exact note event time and a standard deviation in milliseconds.

The KTH rule system was created to model the performance principles of musicians in the context of western classical, jazz, and popular music. A detailed overview of the system by Friberg et al., demonstrates how it is applied to phrasing, micro-timing, patterns, grooves, and many other attributes including performance noise [11]. At a high level, the system takes a nominal score as its input and produces a musical performance on its output as a result of rules whose behaviors are dictated by k values. There have been several practical MIDI implementations of the system that have iterated on rule design and have shown great promise in humanizing an otherwise sterile score at macro and micro temporal levels. In the context of this paper, the simulations of inaccuracies in human motor control are of particular interest. Perception experiments conducted by Juslin et al., have shown that the introduction of a noise rule resulted in higher ratings by listeners in the category of "human" likeness [12].

3. DATA ACQUSITION

In order to understand the nuances of a human percussionist, a system had to be devised to capture raw video in multiple dimensions with sufficient spatial resolution that was synchronized with audio and vibration data [3]. Further, it was critically important to avoid encumbering the musician and instrument with sensors and/or other equipment that could potentially influence the performance [13]. With this in mind, a specific configuration and process was created to capture and extract striking implement tip motion along with pertinent audio and vibration data. The photograph shown in Figure 1 documents the data acquisition system in action during one of several recording sessions. By tilting the drum, providing a calibrated backdrop, and using a single commodity camera running at 240fps with a resolution of 848x480 in portrait mode, a motion tracking application was used to extract the relevant X-Y motion for subsequent analysis.

Figure 1. Recording Session.

4. DATASET

Given the infinite variety of potential percussion performances, a representative subset was needed to establish ground truth. In this context, ground truth refers to a real-world human performance that can serve as a baseline reference for comparative studies.

The Percussive Arts Society has defined an internationally recognized set of rudiments that serves as a "contemporary vocabulary for percussionists" [14]. With a well-documented and defined set of 40 rudiments, a complete dataset was collected using the previously introduced data acquisition system and the participation of a world class percussionist [3, 15]. In an effort to provide a focused analysis, a single rudiment of the set was chosen for the scope of this paper.

The motion data for a performance of the "Double Stroke Open Roll" rudiment can be seen in Figure 2, which is the result of an expert level performance of the score shown in Figure 3. Referring to Figure 2, the green plot is the elevation calibrated Y-Axis motion of the left hand striking implement tip. In contrast, the red plot captures the motion of the right hand striking implement tip. The entire performance spans an excess of three measures at 110 bpm, which amounts to approximately 6.5 seconds in real time along the X-Axis.

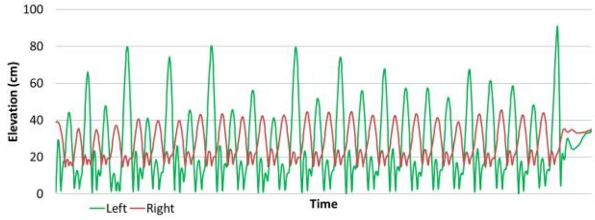

Figure 2. Double stroke open roll motion.

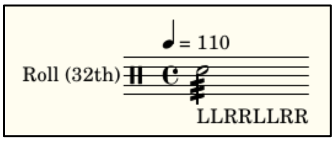

Figure 3. Double stroke open roll score.

The illustration in Figure 4 shows a magnified view of the rudiment along with correlated strike locations. At this level of detail, several triple strikes are visible; however the magnitude of the third strike in each instance is markedly smaller given the prior peak elevation and the reduced discontinuity of the curve at the striking point. From this perspective, the third strike could be construed as gracing the drumhead in preparation for the next intentional striking event.

The elevation amplitude of the left and right striking implement tips in Figure 2 and Figure 4 show a distinct difference that can be attributed to hand dominance [16]. In this example, the left hand is dominant and given its elevation amplitude variation, is far more expressive than the right hand which demonstrates relatively consistent amplitude.

Figure 4. Double stroke open roll motion detail.

Since the drum surface is actually tipped towards the camera at 110°, the absolute elevation can be mapped to a Z-Axis projection that provides a measurement of X-Axis position from the musician's perspective [3]. This component of motion is defined as the position for each striking implement, which has implications with respect to timbre.

5. STOCHASTIC MODEL

A compelling alternative to playing back a prerecorded motion profile in a mechatronic system for a given rudiment is to create a parameter driven model that can create a unique performance on demand. This approach presents an opportunity to identify a set of parameters that when coupled with a stochastic model, results in a unique and "human like" rendering. Given the aforementioned 40 rudiments and the infinite variety of other potential articulations, it is important to create a plug-in model that is extensible and tunable as illustrated in Figure 5. In this conceptual drawing, a collection of models can exist within a system that when triggered, will generate a stream of onset, velocity, and position selection controls based on the current tempo and set of related parameters.

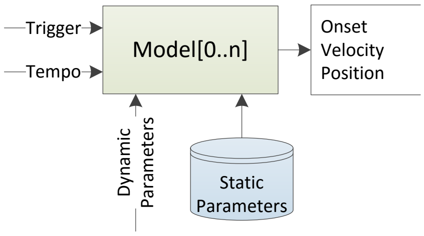

Figure 5. Stochastic model diagram.

In this context, the trigger can be a MIDI "note on" event, an OSC message, or a physical switch. When a particular model is selected as the "patch" and then triggered, the resulting series of onset, velocity, and position values must be subsequently rendered by a robotic or synthesis system. With regards to the former, a mechatronic implementation must faithfully reproduce the striking implement motion that results in the directed sound without introducing significant additional multi-dimensional noise. The combination of actuators, mechanical coupling, and electrical timing will certainly introduce some level of noise, however great care should be taken to minimize this component in order to render an accurate performance.

6. DATA ANALYSIS

By analyzing the onset, position, and velocity of each strike over time for a given rudiment, we can begin to derive a set of parameters for a performance that is composed of stochastic and intentional micro-timing components. As the musician performed the piece, he was interpreting the score in the context of a background click track as heard in monitor headphones. This absolute timing reference provides a means to evaluate each strike in relation to the beat and score. Tendencies to lead or lag the beat can be discovered from both a statistical and temporal perspective.

With drum tilt resulting in a Z-Axis projection, the left and right hand striking implement tip positions can be evaluated over time [3]. Representing the timbre of the strike, this data reveals some of the tonal changes that make a performance unique. As is the case with onset timing, statistical distributions and intention can provide insight into tonal diversity

Lastly, dynamics of a given performance are proportional to the velocity of the striking implement prior to impact. As a result, accents become evident on specific beats, which can be stylistic in nature or traditionally motivated. An example of this would be an emphasis on beat one and three of a four beat measure, which is common in western music.

6.1 Onset

At 110 bpm, a Double Stroke Open Roll rudiment, which is composed of 32nd notes, should have a strike every 68ms. By aligning the first strike of the performance with metrical time zero and measuring the delta of subsequent strikes with expected strikes, we can derive the plot shown in Figure 6. As can be seen from the first set of strikes, the player is attempting to synchronize with the beat; however a progressive lead is evident with a peak of nearly 100ms. As time goes on however we can see the tempo drift towards zero followed by an arc that lags the beat with a peak of 80ms. This artifact provides evidence of a low frequency oscillation in onset timing as the musician chases the beat. The red dotted line in Figure 6 highlights the general shape of the onset drift over three measures.

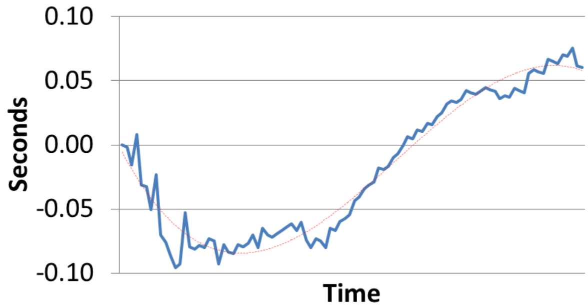

Figure 6. Tempo drift signal.

Incorporating the tempo drift data into a histogram, as shown in Figure 7, reveals a bimodal distribution with a stronger tendency towards leading the beat. Although there are not enough sample points to draw any generalized conclusions about tempo drift, it does illustrate a potential periodic behavior that on the surface seems reasonable.

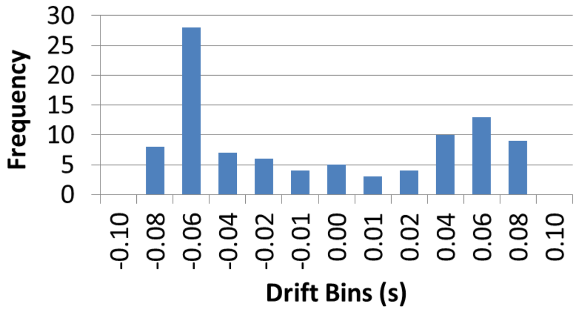

Figure 7. Tempo drift histogram.

By applying Equation 1 over three measures of drift data, the mean tempo delta is -19ms. A standard deviation of 54ms for the same data set was calculated using Equation 2.

$$\mu = \frac{1}{n}\sum_{i=0}^{n} d_i \qquad (1)$$

$$\sigma = \sqrt{\frac{\sum(d_i-\mu)^2}{n}} \qquad (2)$$

In order to preserve the association of the strike parameters it is important to use the average and standard deviation of the same note in each successive bar as opposed to contiguous drift data. A notable side effect of this approach is the decimation of the tempo drift curve as shown in Figure 6. The curve can however be recreated by adding a global sinusoidal oscillation of the appropriate frequency and amplitude to onset events, which for the plot shown in Figure 6, is 0.12Hz with an average peak magnitude of 0.086. It should be noted that vastly different macro-timing characteristics can exist among a population of musicians and even within a series of performances by the same musician.

6.2 Position

The location of impact on a snare drum plays a significant role in timbre. A hit in the middle of the drum has a characteristically lower frequency and shorter dampening time as compared to a strike towards the edge of the drum. This is the result of a complex multi-modal vibration between the drum heads, shell, and snare [17]. Research conducted by Tindale, A. et al., has also shown that striking locations result in distinct timbres that can be classified by training artificial neural networks [18].

A histogram of strike locations shown in Figure 8 reveals a mean of 30.05cm as compared to 43.6cm for the right hand in Figure 9. This separation is intuitive given the Z-axis projection towards the camera since the right hand is at a greater distance. Aside from the relative positions between the left and right striking implement tips, the higher standard deviation of 0.85 for the left hand compared to 0.52 for the right shows that the tonality range is larger on the dominant hand.

Both the left and right hand position histograms have significant skews with value of 0.171 and -0.174 respectively when using Equation 3 on X-Axis motion data. The lack of symmetry strongly suggests an intentional component that coexists with stochastic processes.

$$skew = \frac{n}{(n-1)(n-2)} \sum \left(\frac{d_i - \mu}{\sigma}\right)^3 \qquad \textbf{(3)}$$

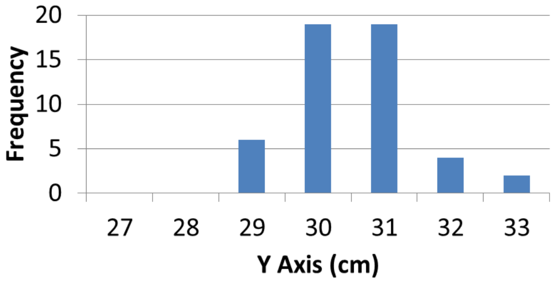

Figure 8. Left hand position histogram.

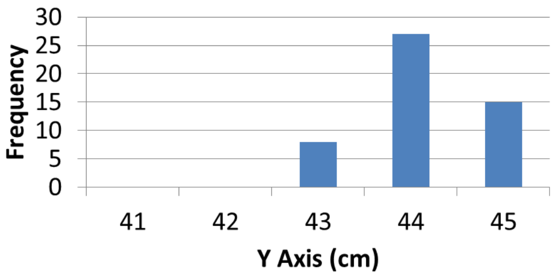

Figure 9. Right hand position histogram.

6.3 Velocity

With respect to dynamics, an average velocity of -600cm/s has the highest incident frequency for the left hand in Figure 10, whereas values near -200 cm/s are more common on the right hand as shown in Figure 11. A statistical analysis of the respective histograms reveals a mean of 844 cm/s and a standard deviation of 240 for the left hand, and a mean of 379 cm/s and standard deviation of 130 for the right hand. This again provides a consistent story that the left hand striking implement is far more expressive than the right hand in terms of amplitude and range.

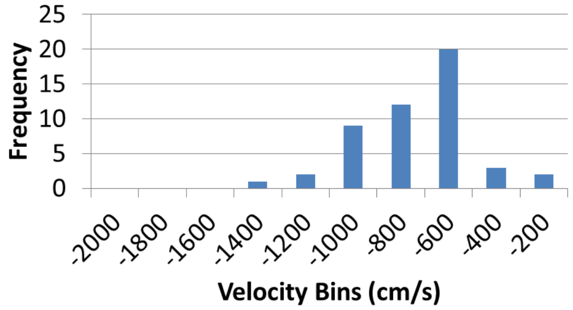

Figure 10. Left hand velocity histogram.

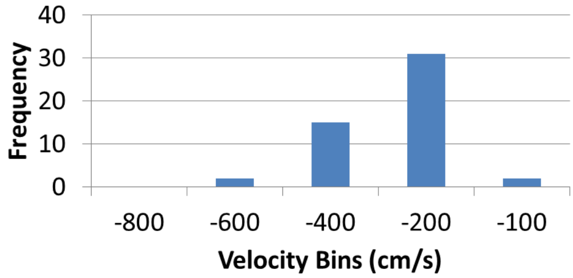

Figure 11. Right hand velocity histogram.

As was the case with the position, the velocity histograms indicate substantial skews. The left hand skew using Equation 3 for the data plotted in Figure 10 is -0.541 and the skew for the right hand velocity histogram shown in Figure 11 is -0.202. As compared to the position histograms, the velocity skewness is much stronger given the emphasis on accented strikes. It is also interesting to note that the left hand skew is higher than that of the right hand. This again correlates with the discovery of a dominate left hand where magnitude, range, and intentional micro-timing are greater.

7. PARAMETER VECTOR

By incorporating the statistical measures for each strike in terms of timing, position, and velocity, we can create a parameter vector that generally describes the rudiment performance. Although a single measure of the Double Stroke Open Roll contains 32 notes, an example of the first eight notes are shown in Table 1. The entries in the parameter vector are used to generate a unique strike using the parameter tuple that consists of onset, velocity, and position.

Equation 4 is used to calculate each onset mean value in the parameter vector as shown in Table 1. Since the rudiment was captured at 110 bpm, a scaling factor can also be applied to the onset time in order to translate it to any desired tempo. The variables of Equation 4 are defined as o_i = expected note onset time, t_r = reference tempo in bpm, t_d = desired tempo in bpm, n = measure aligned note samples, and o_n= measure aligned onset note time.

$$\mu = o_i + \frac{t_r}{t_d}\left[\frac{1}{n}\sum_0^n (o_n - o_i)\right] \qquad \textbf{(4)}$$

The plot in Figure 12 shows the result of computing the onset using a normal distribution with a specified mean and standard deviation for each strike in the measure as directed by the parameter vector depicted in Table 1. The red dotted line indicates the precise metrical position for each note in the score.

Table 1. Parameter vector example.

Note	Onset (μ)	Onset (σ)	Velocity (μ)	Velocity (σ)	Position (μ)	Position (σ)	Hand
1	-0.018	0.004	-1324.101	131.453	29.318	0.161	L
2	0.053	0.004	-1029.712	112.868	29.799	0.053	L
3	0.118	0.004	-515.806	66.100	42.962	0.341	R
4	0.197	0.004	-227.061	80.203	43.666	0.323	R
5	0.252	0.004	-761.860	26.731	30.598	0.510	L
6	0.321	0.004	-620.436	111.393	30.669	0.150	L
7	0.379	0.005	-424.778	136.845	44.004	0.250	R
8	0.457	0.005	-226.775	77.906	44.024	0.099	R

By introducing the concept of a normal distribution derived from a live performance, we can see the resulting effect on the sample implement strike plot in the solid blue line. With respect to the data plotted in Figure 12, Equation 5 can be used to compute each unique onset sample element given the related parameters in the parameter vector given in Table 1. The equation returns an onset sample using the probability density function for a normal distribution as directed by the mean μ and standard deviation σ arguments along with a global drift sinusoid term, which is scaled by the average peak magnitude A. The remaining variables for Equation 5 are defined as o_i = expected note onset time, t_r = reference tempo in bpm, and f_d = drift frequency.

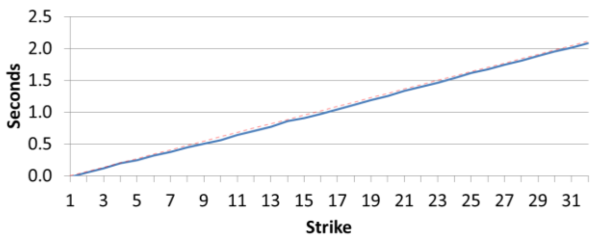

Figure 12 - Onset sample.

$$o_s = N(\mu, \sigma) + \text{A}\sin\left[\frac{o_i t_r f_d \pi}{240}\right] \quad \textbf{(5)}$$

In a similar fashion, the position of each note can be rendered as depicted in Figure 13 by using the normal distribution term of Equation 5 with the related mean and standard deviation parameters in the parameter vector of Table 1.

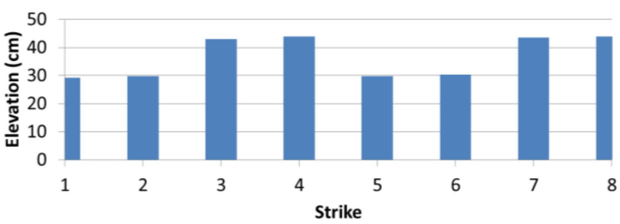

Figure 13. Position sample.

Finally, the velocity including all of the individual accents can be generated as shown in Figure 14, using the normal distribution portion of Equation 5 with the related mean and standard deviation for each note in Table 1. The velocity sample illustrated in Figure 14 also demonstrates the preservation of accent notes, which not surprisingly occur on quarter note strike boundaries {1, 9, 17, 25}.

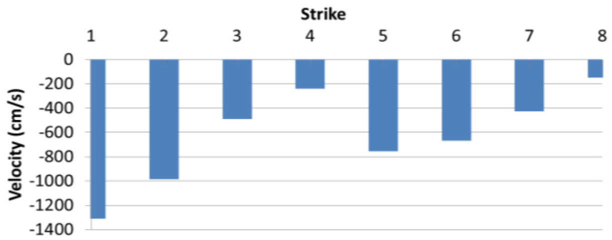

Figure 14. Velocity Sample.

8. EXPERIMENT

A software application was developed to generate N measures of a given rudiment as a MIDI file. For the Double Stroke Open Roll rudiment, the stochastic model used a 32 note parameter vector as depicted in Table 1 with the exclusion of the position/hand parameters, since MIDI "note on" events are restricted to timing and velocity values. The resulting MIDI event timing (horizontal axis) and velocity (height of bar) of the first eight notes for a set variants can be seen in the Sonar X3 DAW screen capture of Figure 15. The top to bottom figure shows four single bar renderings labeled theoretical, dynamics, timing, and timing with dynamics.

When the theoretical performance was played through a snare drum sample, the result was predictably cold and completely devoid of any emotion. In fact, it sounded more like a machine gun! In contrast, when the stochastic model variants were used with the same snare sample, the rendering sounded much more human. By adding pattern and stochastic elements to the velocity in addition to timing, the performance came to life in a way that clearly demonstrated the value of a model that uses multiple dimensions, and whose basic parameters were derived from a real performance. The reader is encouraged to listen to the original recording and renderings by visiting the companion website at http://tinyurl.com/gu2wwlx.

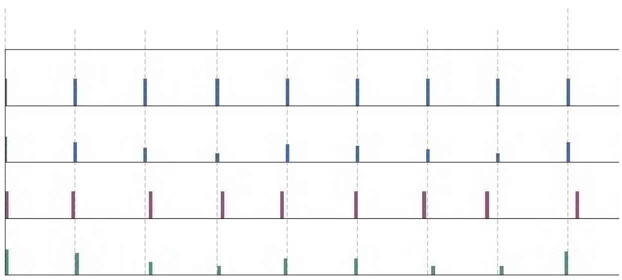

Figure 15. Renderings.

9. CONCLUSION AND FUTURE WORK

Through the novel use of a calibrated data acquisition system, motion extraction, and analysis methodology, we have shown that it is possible to distill a sample performance into a parameter vector that can be used to describe unique human performances. The pragmatic approach to data acquisition enabled the extraction of parameters such as average velocity and position of a strike in an unencumbered manner, which would not have been possible using a typical percussive MIDI trigger. In order to render a convincing robotic performance however, a mechatronic percussionist must be designed and implemented with the appropriate degrees of freedom, spatial resolution, and agility to faithfully reproduce a parameter vector driven stochastic model.

Extending the approach discussed in this paper across the entire set of rudiments and other fundamental percussive articulations will result in a library of parameter vectors that can be used to render unique and life-like performances on demand. Although this represents a daunting manual task, it is possible to automate many, if not all of the steps. With an automated percussion performance parameter vector encoder, it would in fact be trivial to create large parameter vector libraries that could also include multiple percussionists.

Although the concept of a stochastic model for timbre was explored in this paper, an equivalent experiment was not conducted since it cannot be easily mapped to typical MIDI instrument samples. This presents a future opportunity to record snare drum samples that include a timbre dimension with some level of quantization for the striking position. The samples can then be mapped to unique MIDI notes. An additional benefit of this approach will be the inclusion of acoustic and dynamic features that are closely aligned with the original ground truth recordings.

A number of open questions remain that include whether or not micro-timing, dynamics, or position can be generalized for a population of musicians, and if tails and endings have notable features. Answering these questions will require the acquisition and analysis of considerably more motion data, but the outcome will undoubtedly be interesting. In addition, acquiring data in a non-laboratory setting could add further insight into statistical vectors associated with live performances.

10. BIBLIOGRAPHY

[1] M. Davies, G. Madison, P. Silva and F. Gouyon, "The Effect of Microtiming Deviations on the Perception of Groove in Short Rhythms," *Music Perception,* vol. 30, no. 5, pp. 497-510, 2013.

[2] P. Guralnick, Sam Phillips: The Man Who Invented Rock 'n' Roll, Little, Brown and Company, 2015.

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

[4] A. Danielsen, Musical Rhythm in Age of Digital Reproduction, Oslo: Ashgate Publishing Limited, 2010.

[5] J. Fruhauf, R. Kopiez and F. Platz, "Music on the Timing Grid: The Influence of Microtiming on the Perceived Grove Quality of a Simple Drum Pattern Performance," *Musicae Scientiae,* vol. 17, no. 2, pp. 246-2690, 2013.

[6] L. Naveda, F. Gouyon, C. Guedes and M. Leman, "Microtiming Ppatterns and Interactions with Musical Properties in Samba Music," *Journal of New Music Research,* vol. 40, no. 3, pp. 449-460, 2011.

[7] G. Widmer and W. Goebl, "Computational Models of Expressive Music Performance: The State of the Art," *Journal of New Music Research,* vol. 33, no. 3, pp. 203-216, 2004.

[8] S. Dixon, W. Goebl and E. Cambouropoulos, "Perceptual Smoothness of Temp in Expressively Performed Music," *Music Perception,* vol. 23, no. 3, pp. 195-214, 2006.

[9] M. Wright and E. Beradahl, "Towards Machine Learning of Expressive Microtiming in Brazillian Drumming," in *International Compute Music Conference*, New Orleans, 2006.

[10] A. Berndt and T. Hahnel, "Expressive Musical Timing," in *Audio Mostly: 4th Conference on Interaction with Sound - Sound and Emotion*, Glasgow, 2009.

[11] A. Friberg, R. Bresin and J. Sundberg, "Overview of the KTH Rule System for Musical Performance," *Advances in Cognitive Psychology,* vol. 2, no. 2-3, pp. 145-161, 2006.

[12] P. N. Juslin, A. Friberg and R. Bresin, "Toward a Computational Model of Expression in Performance: The GERM Model.," *Musicae Scientiae Special Issue,* pp. 63-122, 2001-2002.

[13] 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.

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

[15] R. Van Rooyen, A. Schloss and G. Tzanetakis, "Snare Drum Motion Capture Dataset," in *New Interfaces for Musical Expression*, Baton Rouge, 2015.

[16] S. Dahl, "Striking Movements: A Survey of Motion Analysis of Percussionists," Acoustic Science and Technology, Denmark, 2012.

[17] T. D. Rossing, I. Bork, H. Zaho and D. O. Fystrom, "Acoustics of Snare Drums," *Acoustical Society of America,* vol. 92, no. 1, p. 84, 1992.

[18] A. R. Tindale, A. Kapur, G. Tzanetakis and I. Fujinaga, "Retrieval of Percussion Gestures Using Timbre," in *International Conference on Music Information Retrieval*, Barcelona, 2004.