Pragmatic Drum Motion Capture System

Robert Van Rooyen
rvanrooy@uvic.ca

George Tzanetakis
gtzan@cs.uvic.ca

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

ABSTRACT

The ability to acquire and analyze a percussion performance in an efficient, affordable, and non-invasive manner has been made possible by a unique composite of off-the-shelf products. Through various methods of calibration and analysis, human motion as imparted on a striking implement can be tracked and correlated with traditional audio data in order to compare performances. Ultimately, conclusions can be drawn that drive pedagogical studies as well as advances in musical robots.

Keywords

Dataset, drum, motion capture, rudiments, robotics

1. INTRODUCTION

Musicians interact with their instruments both in generalized and subtly nuanced ways. The former is part and parcel to learning how to play the instrument given standard instruction in the context of the associated physics. The latter is a fine tuning of the physical interaction that brings out the best musical performance and sound of the instrument. What is the distinction between these two components of a performance? By what methods can we begin to analyze the musician's competence, and by extension, their uniqueness when compared to other performers? To answer these questions in a quantitative manner we need access to real-world data, which in this context is defined as multi-dimensional temporal data from the unencumbered musical performance of a score.

A highly trained and experienced musician can evaluate the quality of a performance purely by ear. This is of course a qualitative and highly subjective measure, but consensus within a population of experts is achievable due to the application of learning constraints [1]. If we breakdown a performance into attributes that can be graded on a scale such as timing, dynamics, and timbre, we can compute the mean, average, and standard deviations for each attribute in a survey. We could further establish a weight based on the experience of the individual evaluator, but ultimately we will derive an informed opinion on the quality of the performance.

In contrast, with access to real-world data as suggested earlier, we can begin to critically evaluate a performance against expected values and in comparison to other musicians. The expected values for attributes can be derived from an original score and can be further adjusted by genre experts with the goal of establishing a reference performance [2]. Although this adjustment can be interpreted as another form of subjective evaluation, the intent is to define ground truth, which will serve as the basis for a subsequent quantitative analysis. The definition of ground truth for our purposes is a dataset that is representative of a quality performance that can be used for comparative studies. A thorough analysis of other performances in reference to ground truth can lead to tangible conclusions about the quality of a given performance and how it is unique within a population of musicians. Further, we will be able to quantify what a high quality performance looks like in addition to what it sounds like. In order to achieve this objective, we established a practical and entirely non-invasive data acquisition method for the recording and interpretation of striking implement tip motion.

2. MOTIVATION

Artistic expression can come in many forms, but a common thread is the notion of individuality. In the case of a musical performance, the uniqueness can be subtle when comparing the work of two highly skilled musicians. Nevertheless, their technique, virtuosity, and style can still set them apart to the trained listener. You might wonder what it takes to become a trained listener. The obvious answer of course is years of formal musical training coupled with an ability to concentrate on subtle sonic detail such as timing, timbre, and dynamics. It comes as no surprise however that computers are also particularly good at differentiating stylistic attributes with the application of expressive models [3, 4].

Extending the concept of automated performance analysis a bit further, it is conceivable that computers can begin to learn what a uniquely human performance encompasses. By comparing musicians against each other and formal notation, statistical patterns and other performance traits such as latency and drift begin to emerge. By applying elements of machine learning, robotic musicians can start to incorporate these nuances into their own renderings, which can dramatically improve the quality of an otherwise sterile, although technically accurate performance.

3. PRIOR WORK

A major challenge of acquiring real-world data is its effect on the system being measured. One can easily postulate that attaching physical sensors to an instrument, musician, or both has the potential to adversely affect the quality of the performance and sound. As an example, previous research conducted by Tindale, A. et al. [5] with percussion instruments have included attaching pressure sensors, contact leads, and accelerometers to sticks and/or drumheads. These sensors inadvertently cause modifications to the sound and perhaps more importantly, the playability of the instrument can be compromised, which can negatively impact the quality of a performance. Moreover, the inclusion of cables and other related hardware can significantly diminish the dynamics of the instrument and the musician's range of motion [6].

The work of S. Dahl, et al. [7] uncovered remarkable detail associated with percussionist motion using high-speed optical motion capture. This approach required the attachment of LED markers on both the subject and the striking implement that worked in concert with the commercially available Selcom Selspot[1] system. Additionally, strain gauges were added to the implements along with an electrically conductive path that provided contact force and duration measurements respectively [8]. As was noted previously, modifications to the striking implements can negatively affect playability. In addition, the cost and complexity associated with this type of motion capture system can be prohibitive. It is important to note however that each approach is motivated by different research questions,

[1] The Selcom Selspot system was first introduced in 1975 and has been used in a wide variety of multi-plane motion capture applications.

which implies that subsequent results can have equally different applications.

In comparison to prior work, our approach has its own unique set of advantages and disadvantages. One of the key benefits is non-invasive data acquisition, which allows the performer to play the instrument in a natural setting without being encumbered with sensors or augmented striking implements. Additionally, the use of low cost commodity equipment provides accessibility to researchers with limited funding and/or access to specialized equipment. As a consequence however, positional accuracy is dictated by the resolution of the camera and the quality of the motion tracking algorithms. Furthermore, manual intervention when extracting motion data due to occlusions or motion blur can result in the introduction of positional errors, which would be manifested as discontinuities or outliers in the data.

4. DATA ACQUSITION

Acquiring real-world performance data in a non-invasive manner does impose limitations on the type of data that can be captured. In some cases however, it is possible to either derive non-measureable values from measureable quantities or infer weighted correlations using calibrated references. For example, a calibrated sound pressure level (SPL) meter can be used at a fixed distance to establish a baseline reference that is synchronized with the audio recording.

4.1 Video

A study conducted at Harvard University by A. Hajian, et al. [9] concluded that the upper bound of the impact rate for a drum roll performed by an accomplished musician is on the order of 30Hz. In this case however, the signal of interest is not the impact rate, but rather the motion that results in the impact rate. To capture the related motion sufficiently we must have a video frame rate that is "high enough" to produce reasonably smooth data [10, 11, 5].

A survey of commercially available video cameras resulted in a number of candidate products that ranged widely in cost, features, and size [10, 2, 7, 12, 13]. The camera we selected was the GoPro HERO3+ Black Edition[2], which is a very versatile camera, intended for rugged outdoor use when contained within its protective housing. The GoPro supports a variety of resolutions and frame rates that includes 848x480 at 240 frames per second. By rotating the wide angle field of view by 90° the relatively inexpensive camera can produce the desired quality and sampling rate for the motion capture system.

The video data acquisition elements are composed of the snare drum, foam board backdrop, video camera, and key light as shown in Figure 1. As depicted in the diagram, the drumhead is tilted at an obtuse angle θ of 110°, which results in a Z-axis projection that enables depth perception. This is an important attribute of the configuration as it eliminates the need for a second camera in order to resolve drumhead region mapping. The video camera has a wide angle lens and is used in portrait mode in order to utilize the full 848 vertical lines of resolution, which results in a field of view that encompasses a typical range of motion for a performance. Finally, an inexpensive project lamp outfitted with a 100W compact florescent bulb is used as the key light [2].

[2] Additional information on the GoPro HERO3+ Black Edition can be found at http://bit.ly/1zVt7DP.

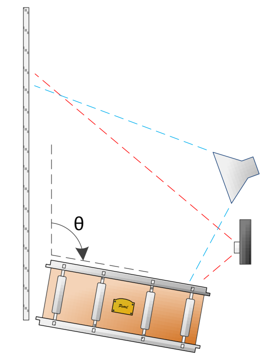

Figure 1. Camera orientation

An example of the actual view from the camera's perspective can be seen in Figure 2. In this frame grab, the striking elements, drum surface, and backdrop are visible. In addition, the backdrop contains a dimensional reference placard that enables video tracking software distance calibration and radial distortion compensation.

Figure 2. Camera view

Each tip of the striking implements has been coated with an enamel paint, which in this case is green for the left stick, and red for the right stick. These colors are used by the tracking software as unique contrast to distinguish individual motion [10, 2, 12]. The only caveat is that the red tip must not occlude the green tip, which can be accomplished by using a slight horizontal offset when performing, or potentially adjusting the lateral camera position with angle compensation using calibration references.

The software [2] that was used to extract tip motion is the "Tracker" application[3] for Windows, which is a project of the Open Source Physics organization. This program provides a rich set of tools for importing content, filtering video, calibrating distance, tracking motion, and extracting a variety of time referenced values.

4.2 Audio

Audio recording is naturally at the very heart of capturing a musical performance. In this instance, an industry standard Shure SM57 dynamic microphone[4] was used along with an Edirol UA-25 (24-bit/96kHz) audio interface[5], which includes an integrated dual channel preamp. The audio interface was connected to a high-performance laptop running Windows 7

[3] Detailed information or the cross-platform Tracker v4.85 application, © 2014 Douglas Brown, is available at http://bit.ly/1qWLHem.

[4] Additional information on the Shure SM57 dynamic microphone can be found at http://bit.ly/1uCQMbC.

[5] Detailed specifications for the Edirol UA-25 audio interface are located at http://bit.ly/1s2yRNL.

that was configured with the Sonar X1 Digital Audio Workstation[6] software.

4.3 Vibration

In contrast to capturing variations in sound pressure through the air, recording the vibration of the drumhead can reveal other subtleties related to a given performance, such as timbre and tuning characteristics [5]. The transducer used in our system was the Roland RT-10S acoustic drum trigger[7]. Drum triggers are normally used with an electronic "brain" that interprets the analog signal in order to determine MIDI onset and velocity data for sound samples while filtering out false triggers. In this configuration however, we can benefit from the raw analog signal coming directly from the transducer, which is mechanically connected to the drumhead surface. A spectral analysis of the signal data can reveal the resonant frequency of the drum along with key harmonics that are directly influenced by the mechanical composition of the drum. The signal from the acoustic drum trigger is treated as a standard microphone input.

5. DATA PROCESSING

5.1 Session Calibration

Calibrating the data acquisition system is a critical step in understanding and categorizing performance data [5]. Aside from validating equipment configuration and position specifications, two of the key calibration elements are timbre and dynamics.

5.1.1 Timbre

A component of snare drum timbre is dependent on the location of drumhead impact. A performance of the score depicted in Figure 3 with motion capture enables the quantization of impact locality.

Figure 3. Timbre score

As illustrated in Figure 1, the obtuse angle of the drumhead in relation to the camera field of view results in a Z-axis projection. This projection is quantized into three regions, which are labeled as center, one-third, and rim [13] as illustrated in Figure 4.

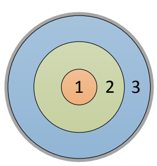

Figure 4. Impact regions

To identify these regions, we must process the Y-axis data from the calibration performance by subtracting the global minimum as defined in Equations 1, 2.

$$c = \min(Y) \quad (1)$$

$$z(t) = y_t - c \quad (2)$$

This is followed by defining static thresholds a and b that ideally delimit the impact regions. The point of impact represents a local minimum, which can be simplistically determined by evaluating a given sample with adjacent values that meet a minimum distance criterion m as depicted in Equation 3[8]. Identification is achieved by locating the point of impact and mapping the elevation value to a region based on the established thresholds as shown in Equation 4. The plot in Figure 5 demonstrates detected hits from a performance of the score in Figure 3 and how they were categorized as center (C), one-third (O), and rim (R).

$$h(n,t) = \begin{cases} z_t, z_t < z_{t-1} - m \wedge z_t < z_{t+1} - m \\ undefined \end{cases} \quad (3)$$

$$c(n) = \begin{cases} 'C', h_n < a \\ 'O', a < h_n < b \\ 'R', h_n > b \end{cases} \quad (4)$$

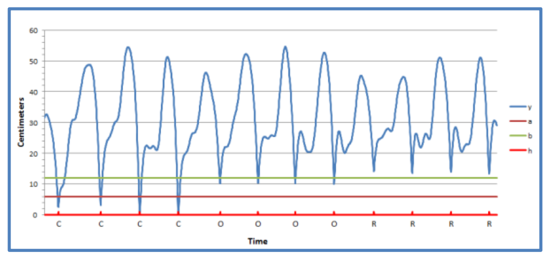

Figure 5. Timbre calibration

5.1.2 Dynamics

Calibration of dynamics is an attempt to identify a range of sound pressure levels (SPL), from very quiet to very loud. The score in Figure 6 defines a performance that is a two measure crescendo from piano pianissimo to forte fortissimo, for which the related audio signal from a performance can be seen in Figure 7.

Figure 6. Dynamics score

As depicted in Figure 7, the peak and hold function of Equation 5 captures the successive dynamic level increase of each impact [7]. The regression curve was derived in MATLAB using a second order polynomial as shown in Equation 6 with coefficients {-0.0286, 0.3542, -0.1047}.

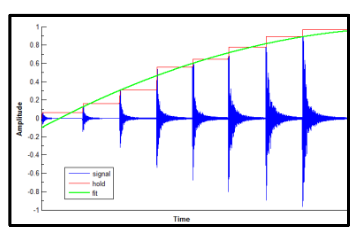

Figure 7. Dynamics curve

---

[6] The Sonar X1 Digital Audio Workstation software is a product of Cakewalk, which produces a variety of audio production tools: http://bit.ly/1wlfPjy.

[7] Information on the Roland RT-10S acoustic drum trigger can be found at http://bit.ly/1ri2rMV.

[8] A more robust min/max search algorithm for multivariate data should be used in practice.

$$h(x,t) = \begin{cases} x_t, h_{t-1} < x_t \\ h_{t-1} \end{cases} \quad \textbf{(5)}$$

$$f(x) = c_1x^2 + c_2x + c_3 \quad \textbf{(6)}$$

5.2 Onset Time

In addition to many other attributes, it is possible to detect the onset time for each strike as shown in Figure 8 using transient selection mode in Sonar X1. The recording of eighth note strikes (red onset) illustrates instances of both leading and lagging the reference beat (vertical black bands) at 96bpm.

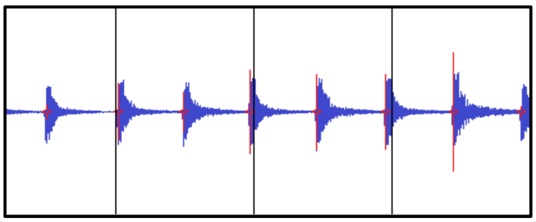

Figure 8. Onset time

By examining the delta between actual onset time and meter, statistical metrics can be computed for the performance. In this case, the median delta is -33ms, which indicates a bias towards leading the beat. The tendency towards leading is clearly visible in Figure 8 when comparing onsets to the established meter.

5.3 Spectral Plot

The result of plotting a power spectrum for both the vibration transducer and dynamic microphone can be seen in Figure 9. Despite a common source, the spectral profile is quite different given wave propagation through air versus a solid material, where the latter is a complex combination of the transducer location, drumhead, and all of the physical components of the drum.

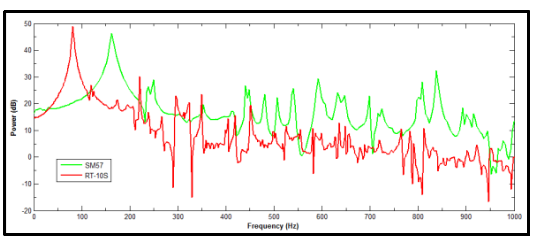

Figure 9. Power spectrum

6. CONCLUSION AND FUTURE WORK

Capturing the motion associated with percussion instrument performances requires an infrastructure of equipment, software, analytical processes, skilled musicians, and a formal calibration technique. The result however is the ability to evaluate a performance from multiple vantage points in a non-invasive manner with respect to the musician and the instrument. It is also evident that meaningful motion capture can be accomplished with minimal resources and at a relatively low cost.

A specific area of research that can greatly benefit from validated ground truth and a repeatable acquisition system is the development of robotic musicians. Many of the implementations to date have focused on absolute timing accuracy and speeds that exceed human ability. In contrast, designing algorithms and electromechanical systems that can learn to reproduce a truly human inspired performance could represent a new and exciting branch to an existing body of work.

7. BIBLIOGRAPHY

[1] K. A. Ericsson and C. A. Lehmann, "EXPERT AND EXCEPTIONAL PERFORMANCE: Evidence of Maximal Adaptation to Task Contraints," *Annual Review of Psychology*, vol. 47, pp. 273-305, 1996.

[2] M. Miura, "Inter-Player Variability of a Roll Performance on a Snare-Drum Performance," in *International Conference of the Forum Acusticum*, Budapest, 2005.

[3] E. Stamatatos and G. Widmer, "Automatic identification of music performers with learning ensembles," *Artificial Intelligence*, vol. 165, pp. 35-56, 2005.

[4] 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.

[5] 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.

[6] M. Collicutt, C. Casciato and M. M. Wanderley, "From Real to Virtual: A Comparison of Input Devices for Percussion Tasks," in *International Conference of New Interfaces for Musical Expression*, Pittsburgh, 2009.

[7] 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.

[8] S. Dahl and E. Altenmuller, "Motor Control in Drumming: Influence of movement pattern on contact force and sound characteristics," in *International Proceedings of Acoustics*, Paris, 2008.

[9] A. Z. Hajian, D. S. Sanchez and R. D. Howe, "Drum Roll: Increasing Bandwidth Through Passive Impedance Modulation," in *International Conference on Robotics and Automation*, Albuquerque, 1997.

[10] E. Schoonderwaldt, N. Rasamimanana and F. Bevilacqua, "Combining Accelerometer and Video Camera: Reconstruction of Bow Velocity Profiles," in *International Conference of New Interfaces for Musical Expression*, Paris, 2006.

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

[12] H. Kawakami, Y. Mito, R. Watanuma and M. Marumo, "Analysis of Drum Player's Motion," in *International Proceedings of Acoustics*, Paris, 2008.

[13] A. Bouenard, S. Gibet and M. M. Wanderley, "Enhancing the Visualization of Percussion Gestures by Virtual Character Animation," in *International Conference of New Interfaces for Musical Expression*, Genova, 2008.