Hearing Movement: How Taiko Can Inform Automatic Recognition of Expressive Movement Qualities

Shannon Cuykendall [1]
scuykend@sfu.ca

Michael Junokas[2]
junokas@illinois.edu

Mohammad Amanzadeh [2]
amanzad2@illinois.edu

David Kim Tcheng,[2]
dtcheng@illinois.edu

Yawen Wang
asianrainwerk@mac.com

Thecla Schiphorst [1]
thecla@sfu.ca

Guy Garnett[2]
garnett@illinois.edu

Philippe Pasquier [1]
pasquier@sfu.ca

Simon Fraser University[1]
Surrey, BC Canada

University of Illinois [2]
Urbana-Champaign, IL United States

ABSTRACT

We describe the first stages of exploratory research undertaken to analyze expressive movement qualities of taiko performance, a Japaense artistic practice that combines stylized movement with drumming technique. The eventual goals of this research are to answer 1) Can expressive visual qualities of taiko be heard in the sound and 2) Can expressive sonic qualities of taiko be seen in the movement? We achieved high accuracy across multiple machine-learning algorithms in recognizing key sonic and visual qualities of taiko performance. In contrast to many current methods of studying expressive qualities of movement, we inform our data collection process and annotations with taiko technique. We seek to understand how the fundamentals of taiko create expression. More broadly, we suggest that codified artistic practices, like taiko, can inform automatic recognition and generation of expressive movement qualities that have been challenging to reliably classify, parse, and detect. In future work we propose ways to generalize expressive features of taiko so they can be recognized in other movement contexts.

Author Keywords

Taiko; Musical Gesture; Sound-Producing Gesture; Expressive Movement; Machine Learning; Movement Classification

ACM Classification Keywords

H.1.2 User/Machine Systems: Human Factors; J.5 Arts and Humanities: Performing arts (e.g., dance, music)

INTRODUCTION

The need to recognize and engage with the subtleties of human movement has become an important goal in third wave HCI. Interaction designers now seek to create richer interactions that respond not only to *what* action is performed, but also to *how* an action is performed. Many researchers have turned to dance and movement practices to inform how to capture and extract these expressive qualities from movement data.

We describe exploratory research undertaken by researchers from Simon Fraser University, the University of Illinois, and an independent taiko artist. Taiko is a Japanese artistic practice that integrates stylized choreographed movement with drumming technique. We propose that taiko technique can offer valuable insights into recognizing expressive qualities of movement because of its close interconnection between sonic and movement qualities, allowing us to analyze expressive features from two modalities. The eventual goals of this research are to answer the following: 1) Can expressive visual qualities of taiko be heard in the sound? and 2) Can expressive sonic qualities of taiko be seen in the movement? We address the first steps in answering these questions by illustrating that we can produce accurate machine-learning models that are capable of recognizing key aspects of taiko performance in visual and auditory modalities. These findings can be applied to interactive performances and pedagogical systems for taiko performance. In future work we discuss how expressive taiko features can be generalized and recognized in other movement contexts.

BACKGROUND

Expressive qualities are broadly defined in literature. According to Camurri et al (2004), "Expressive content concerns aspects related to feelings, moods, affect, intensity of emotional experience" and is not necessarily tied to a denotative meaning. This definition implies that expressive qualities can be found in any movement that contains information about the state of the mover—from everyday actions, such as walking, to actions executed in performance settings [7].

In Laban Movement Analysis (LMA), a prominent movement classification system (also referred to as Laban Bartenieff Movement Studies [26]), expressive qualities of movement are articulated through four main categories of movement: Body, Effort, Shape, and Space. Generally, the Body category refers to what part of the body is moving and the initiation of a movement. The Effort category specifies how an action is executed relating to the dynamics, qualities, energy and inner intent of the movement. The Shape category describes the relationship of the mover to the environment, to others, and to one's self . Lastly, the Space category refers to the level, direction, pathways, and the extension of the limbs

MOCO'15, August 14–15, 2014, Vancouver, BC, Canada.

ACM ISBN 978-1-4503-3457-0/15/08...$15.00
DOI: **http://dx.doi.org/10.1145/2790994.2791004**

from the body's core. While Effort may appear to be the category most linked with human expression, in actuality these categories of movement overlap and inform one another in numerous ways. As certified movement analyst Karen Studd points out, "...the functionality and expressiveness of movement is intertwined" [26]. This suggests that expression can arise from any dimension of movement.

Researchers have followed two main routes to annotate expressive movement: 1) Comparing basic human emotions or affect with kinematic measures (e.g. [23, 4]) and 2) The use of movement classification systems, such as LMA (e.g. [21, 2]). Often these techniques are used in combination with one another, for example mapping movement qualities defined in LMA to emotional states (e.g. [19, 14]). While these methods have had some success, many challenges in annotating movement data still remain. For example, studies have suggested that culture [18] and one's visual and physical familiarity with movement [6] may influence how one perceives movement, making it difficult to claim that movement labels are universal. Furthermore, there are many proposed methods of segmenting continuous motion for analysis (e.g. [8, 27, 25, 20]); however, there is generally no accepted way to determine the length or boundaries to best analyze continuous motion.

In contrast to this research, we instead start from a more narrow vocabulary that is informed by a *codified artistic practice*. Codified artistic practices refer to techniques that have been developed over many years and have established "correct" and "incorrect" form. While many researchers have used codified movement practices, such as ballet [3] to study expressive movement, the labels applied to the data are often external to the technique. We instead seek to understand how the fundamentals of the technique can create expression.

In using codified movement practices to inform data collection procedures we can overcome many of the challenges in analysis, such as movement segmentation and establishing ground truth annotations. For example, in taiko each action can be mapped to a discrete sound, allowing for accurate and easy segmentation. Codified practices are also culturally tied to the time and place they were developed and practitioners of the technique are trained to perceive qualities in a similar manner. Therefore, codified techniques, like taiko, serve as living records of perceptually evident expressive qualities within their specific culture. Understanding how expressive movement is perceived in different cultures can eventually help inform more universal labels.

Furthermore, we propose that the specific qualitative groupings developed within codified artistic practices may be important to the recognition of expressive qualities. As stated earlier, the LMA system proposes that expression can arise from multiple dimensions of movement. However, while any combination of movement qualities are conceivably possible in LMA, not all of them are necessarily perceptually evident. Therefore, a more narrow classification system, like taiko, can help inform the dimensions of movement to explore. Taiko is an ideal practice to study expressive features because sound and movement are equally emphasized and integrated in the technique, allowing one to study expressive features through two modalities.

The integration of sound and movement has been extensively studied through three main approaches: 1) examining the musical gestures of a performing musician 2) examining the musical gestures of listeners perceiving the sound and 3) examining the sonifications of movers. The motivation behind much of this research is for sound synthesis and the design of gestural musical interfaces. Most relevant to our research are studies concerning "sound-producing musical gestures" which are movements that are directly responsible for creating a sound such as striking a drum or piano key [15]. A few key works specifically related to our research include, the AoBachi drumstick interface, which was also informed by taiko technique [29], a classification framework for analyzing tympani drum-strokes described in Bouënard et al (2010) [5], and multiple studies on sound perception and drum-stroke motions (e.g. [10, 11, 12]).

OVERVIEW OF TAIKO

Taiko, meaning "fat drum," most commonly refers to a Japanese form of ensemble drumming that became popularized as a performing art in the mid-20th century. However, the roots of taiko are over a thousand years-old and began in villages primarily to communicate over long distances and to build a sense of community. Taiko has been referred to as the "heartbeat of Japan" and is seen by many as not just a technique, but as a spiritual path [28].

Sound	Description	Time
Don	1 loud beat	♩
Doro	2 loud beats	♫
Tsu	1 soft beat	♩
Tsuku	2 soft beats	♫
Ka	1 rim beat	♩
Kara	2 rim beats	♫
Su	rest	♩

Table 1: *Kuchi-showa*, phonetic vocal notation of taiko and associated rhythms

Taiko is traditionally passed down through *kuchi-showa*, a phonetic vocal notation system that includes the following sounds: *Don*, *Doro*, *Tsu*, *Tsuku*, *Ka*, *Kara*, and *Su*. Don is a loud sound, Tsu is a soft sound, Ka is a rim sound and Su is a rest. Doro, Tsuku and Kara refer to either 2 Don, Tsu, or Ka beats (respectively) played in succession (See Table 1). Students often learn rhythms by singing the sounds before playing them on the drum.

In taiko, it is critical that both sound and movement are practiced and perfected. Beginning-level students work on mastering fundamentals related to both sound and movement such as correct stance, the trajectory of the arms, grip of the drumstick (called *bachi*), rhythm and distinguishing between Don and Tsu sounds.

Since much of taiko is passed down through vocal notation, taiko training differs slightly from region to region. While the fundamentals of taiko remain for the most part consistent, teachers and taiko artists have created individual ways of articulating and notating them. Some taiko artists teach primarily through demonstration with very little verbal feedback while others have used western musical notation and terminology to describe taiko technique (See Figure 1).

Jason Overy, of Vancouver's Uzume Taiko [22], created a taiko syllabus that distinguishes between many aspects of taiko that are often demonstrated, but not always verbally articulated. In Overy's syllabus specific body movements are defined and linked to taiko sounds. All taiko sounds can be performed within five *gears* that refer to the body part or action initiating the sound, such as fingers, wrist, elbow, full arm, and jumping. Within each gear there are five *levels* that refer to the angle of the drumstick in relation to the drum. Level 1 is 0 degrees and level 5 is 90 degrees. This level system is inspired by techniques used in drum corps training (See Figure 2).

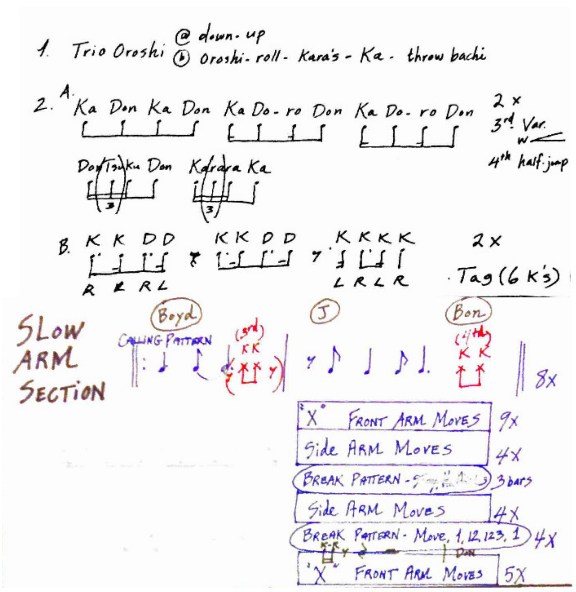

Figure 1: An example of a taiko score influenced by western music notation. ©Yawen Wang, 2004

TAIKO INFORMED DATA COLLECTION

We used Overy's syllabus to inform our data collection process. We captured a variety of taiko drum-strokes performed by Yawen Wang, an expert taiko performer who is also a coauthor of this paper. Wang has trained professionally for over five-years and performed with Uzume Taiko, a prominent taiko group in Vancouver, BC. In addition to taiko training, she also plays a variety of other instruments including hand drums, the theremin, and the piano. She has also worked as a dance accompanist and has trained in various dance styles.

We recorded Don and Tsu strokes played in similar gears and levels, including Gear 2 Level 4, Gear 3 Level 4, and Gear 4 Level 5. We chose to start with these particular combinations because they felt "natural" to the performer when playing

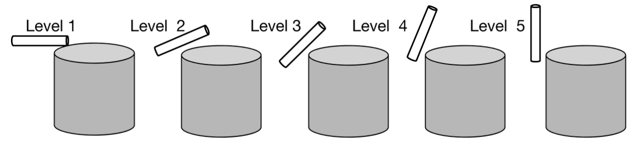

Figure 2: Taiko drumstick levels as defined by Jason Overy of Uzume Taiko in Vancouver, BC Canada.
©Shannon Cuykendall, 2015

both Don and Tsu sounds. Furthermore, each recorded stroke was played as either *legato*, *staccato*, or *marcato* (See Table 2). Legato refers to smooth, connected phrasing. Staccato describes shortened disconnected sounds, and marcato is an accented or forceful sound. Only Don sounds can be played with a marcato quality and only Tsu sounds can be played with a staccato quality. Each drum-stroke was recorded with both left and right arms, but for the purposes of this paper we analyze only the right-handed strokes. For each recording the performer played the stroke multiple times in a row, taking a rest between each stroke so that the sound completely diminished before starting the next stroke. While drumming is often more continuous in performance contexts as Dahl (2011) suggests [11], we chose to record the strokes this way so we could analyze the discrete actions and sounds that are fundamental to taiko technique.

Gear	Level	Sound	Articulation
1 (fingers)	1	Don	Marcato
2 (wrist)	2		Legato
3 (elbow)	3		
4 (full arm)	4	Tsu	Staccato
5 (jump)	5		Legato

Table 2: Each stroke we recorded had four labels: gear, level, sound and musical articulation. The gears, levels, and sounds can be combined in any manner. Tsu can only be played with a staccato or legato quality, while Don can only be played with a marcato or legato quality.

DATA CAPTURE SETUP

Motion Capture

In order to digitally capture movement, we used Microsoft Kinect sensors. The Kinect captures movement through the generation of depth maps, utilizing camera and infrared sensor data. From the depth maps a variety of informative data can be extrapolated, such as a skeleton frame representing positional data of the subject. While the Kinect is lower resolution than other motion capture systems, it also provides subjects with a minimally invasive situation where they are not required to wear any additional devices that may inhibit their ability to replicate the physical qualities of performance. This is especially important in the case of taiko, where expressive performance demands high levels of control over a diverse range of nuanced percussion techniques. Using the Kinect data we were able to effectively recognize key aspects of taiko performance, demonstrating little need to use higher resolution motion capture technology at this stage in our research.

Audio Capture

We captured audio using cardioid condenser microphones in three positions in order to fully realize the expressive sounds produced by the taiko drum and performer. The first microphone was placed inside the open base of the drum in order to capture lower frequency, resonant sounds. The second microphone was placed above the skin of the drumhead in order to capture higher frequency, attack sounds. The third microphone was placed on the right-side of the drum, at the performer's head-level, in order to capture vocalizations and the ambient room sound.

The three channels of mono microphone analog audio data were converted to digital audio through an analog-to-digital audio interface sampling at 44.1KHz. Once the audio data was digital, it was minimally processed in mixing software in order to maintain the sonic integrity of the analog signal.

CLASSIFICATION OF TAIKO TECHNIQUE

Using machine-learning techniques, we performed various experiments to classify key aspects of taiko technique. To classify the taiko sound data we used two different methods: a Naïve Bayes (NB) classifier [16], and an Instance Based (IB) method with parameter optimization [1]. We also used Hidden Markov Models (HMM) [24] for both sound and visual data. Since the results of the HMM sound classifier were similar to the results of the other two techniques (NB and IB), we only report the HMM results for the visual data.

In this work we explored three properties of taiko technique: a 2 class test between Don and Tsu techniques, a 2 class test between staccato/marcato and legato musical articulations, and a 3 class test to distinguish the performance in different gear/level combinations, including gear2/level4, gear3/level4, and gear4/level5 (See Table 3).

Taiko Sound Classification

Machine-Learning Algorithms

We classified sound primarily using a NB classifier, but also conducted a few additional tests using an IB algorithm.

NB creates a fixed-sized model with weights for every feature. NB models are simple, easily understandable, and have extremely fast training and recognition. They assume a simple monotonic relationship between feature values and class probabilities and there are no important feature interactions.

In contrast to NB, IB algorithms do not create a fixed model, but instead memorize all examples at maximum resolution and use these memorized *instances* to make predictions. The final class prediction for an example is determined by finding its nearest neighbor in the instances that have been stored by the algorithm. In our IB algorithm the distance between the examples is measured by a city block distance metric. IB can be slow for classification due to N^2 complexity where N is the number of training examples. The advantage of IB is that it has the representational flexibility to learn any classification function, provided enough training examples are available. Our IB algorithm utilizes an optimization algorithm to systematically search for good system parameters for the given problem. For this experiment, the optimizer adjusted 6 parameters affecting example representation: *SampleTrim*, *DampingRatio*, *MinBandFrequency*, *MaxBandFrequency*, *SpectraSampleRate*, and *SpectraNumBands*.

Audio Sample Representation

We considered multiple sample representations for machine learning, varying in duration and focus as well as spectral extraction method.

Through empirical evaluation, we found that all drum-stroke sounds decayed in less that 2 seconds. It was unknown whether the whole sound event should be used for sample generation or if samples should focus primarily on the attack portion of the sound. Thus, we considered both long duration sound samples (the full 2 seconds) and short duration sound samples (focusing on the attack). We defined the attack portion of the strike as the onset of the attack (just preceding 100% energy) until 30% reduction of amplitude (70% total energy) was achieved. Through empirical evaluation, all attacks lasted no more than 1500 audio samples. Thus, the segmented samples began at the onset of the attack and continued until 2048 samples after (0.05s).

We used two different methods to compute spectrograms from the sound files: FFT and filter banks. For FFT we used 2 sample sizes. For the longer samples, we used a window size of 512 audio samples with an overlap of 256 (initial test). For the shorter attack focused samples, we used a single 2048 window with no overlap (secondary test). For both tests, we subtracted a noise profile spectrogram, generated from audio of the room in silence. Additionally, for the secondary test, we normalized each spectrogram within the frequency domain. For filter banks, we used up to 580 band pass filters and energy detectors to generate the final spectrogram. Each energy detector is like a tuning fork which sympathetically vibrates when exposed to specific sound frequencies like the hair cells in the inner ear.

Performance Evaluation

Machine-learning algorithm performance was measured using leave-one-out cross-validation (LOOCV). This method takes one sample from the dataset, removes it, trains the classifier with all other samples, and then tests the removed sample. This process was repeated for all samples and an overall accuracy was obtained by averaging all results (See Table 3).

Taiko Movement Classification

Machine-Learning Algorithm

We used a HMM to classify the taiko visual data. Since HMM parameters have a critical role in a model's performance, we identified the optimal values for each test. The most important parameters of HMMs are the number of states, the number of Gaussian mixtures required for modeling each state, and the architecture of the states. The usual architectures are ergodic and left-right (LR). With an ergodic architecture, each state can be reached from all other states, while in a left-right architecture, each state can only be reached from states with a lower order.

We experimented with various kinematic features of motion data to build the HMMs, and modified the HMM parameters to obtain the maximum accuracy. To choose the optimal parameters, we also considered the computation time. Since the LR architecture requires less computation, it is more suitable for real-time purposes. LR architecture is also convenient for the type of data in these experiments, due to the similar overall shape of the striking gestures (starting with raising the drumstick, attacking the drum, and the final rebound). We used LR architecture for all musical articulation tests, however we found that the ergodic architecture worked better for gear/level and Don/Tsu tests. We also found that 4 or 6 hidden states worked best in all the classifiers.

Movement Sample Representation

To explore visual features of taiko movement we used the Kinect skeleton data, extracting the X, Y, Z position of the wrist-joint of the performer. The samples were manually segmented to include the entire length of each stroke, starting from raising the drumstick and ending up to 10 frames after the stroke (see Figure 3).

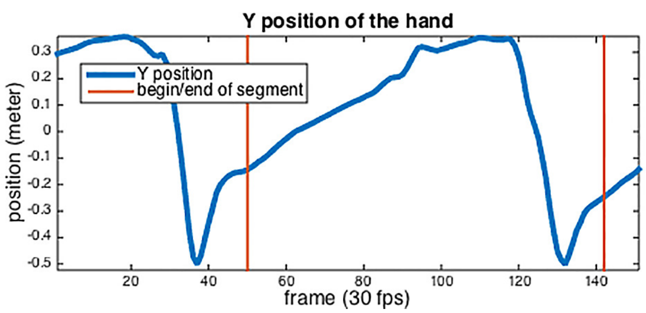

Figure 3: Manual segmentation of the drum-stroke using the vertical position of the wrist-joint.

$$\mathbf{V}_t = \frac{\mathbf{P}_t - \mathbf{P}_{t-1}}{\Delta t} \qquad \text{velocity} \qquad (1)$$

$$\mathbf{A}_t = \frac{\mathbf{V}_t - \mathbf{V}_{t-1}}{\Delta t} \qquad \text{acceleration}$$

$$S_t = \|\mathbf{V}_t\| \qquad \text{speed}$$

$$\theta_t = \cos^{-1}\left(\frac{\mathbf{V}_t \cdot \mathbf{V}_{t-1}}{\|\mathbf{V}_t\| \, \|\mathbf{V}_{t-1}\|}\right) \qquad \text{angle}$$

We generated 5 different features from the joint positions. Velocity and acceleration were computed by taking the first and second derivative of movement position over time. We also computed the dot product, cross product, and the angle between consecutive displacements of the wrist-joint's position to capture more information about the curvature of movement (see Equation 1).

Performance Evaluation

We trained one HMM for each class of taiko drum-strokes. To identify the correct class of a sample, we compared the likelihood of the sample using each model, and chose the model which reported the highest likelihood. In order to avoid locally optimal solutions, the HMMs were trained 10 times and the model with the highest likelihood was chosen. We tested the model with 20 samples for each gear/level combination, and used a 3-fold cross validation, with 70% of data for training, and 30% for testing. Our models were able to classify taiko drum-strokes with 97% to 100% accuracy in all four experiments (See Table 3).

RESULTS AND DISCUSSION

Tests	Gear/Level	Audio1 (NB)	Audio2 (NB)	Audio1 (IB)	Visual (HMM)
Gear/	2/4	0.83	0.66	na	0.92
Level	3/4	0.79	0.48	na	0.92
	4/5	0.82	0.69	na	1.00
Don/	2/4	1.00	1.00	na	0.97
Tsu	3/4	1.00	1.00	na	0.97
	4/5	1.00	1.00	0.99	1.00
DonMar/	2/4	0.97	0.95	na	0.97
DonLeg	3/4	0.77	0.84	na	1.00
	4/5	0.83	0.78	0.98	1.00
TsuSta/	2/4	0.66	0.59	na	0.97
TsuLeg	3/4	0.50	0.53	na	1.00
	4/5	0.79	0.55	0.98	1.00

Table 3: Results of audio and visual analysis of gear/level, Don/Tsu, and musical articulations. The number of samples used for each test varied: Gear/Level Audio and Visual (120 samples), Audio 1 and 2 NB (81 samples for 2/4, 82 samples for 3/4, and 79 samples for 4/5), Audio 1 IB (40 samples in each class), Visual HMM (20 samples in each class). Audio1 = 2s length (not normalized), Audio2 = 0.045s length (spectrally normalized), and visual samples are manually segmented (starting from raising the drumstick and ending up to 10 frames after the stroke.)

Audio Analysis

A relatively high accuracy was achieved in our initial (Audio 1 NB) analysis of gear/level. The discrepancy in accuracy could be seen as a lack of independence between classes for the tests, a difficulty seen in creating effective physical models for bowed string instruments [13], percussion [9], and other instruments [17]. In the Audio 2 (NB) test, we minimized the effect of amplitude on the data features by additionally reducing the sample size and normalizing the FFT. This led to lower accuracy in recognizing the gear/level of a stroke. We can interpret this to mean that the gear/level of a stroke may rely more on amplitude as a discriminating feature. In similar research, Dahl (2004) found that dynamic level was correlated to the preparatory height of the stroke [10].

Both NB and IB classifiers were able to distinguish between Don and Tsu sounds with a high degree of accuracy. To investigate this more thoroughly, we decided to apply principal component analysis to the spectrograms of data in these two classes and visualize the output. The first two principle components demonstrated a clear distinction between Don and Tsu, with Don having a more energy in the first component (See Figure 4). This suggests that Don and Tsu sounds vary both in timbre and amplitude. In taiko technique, Don and Tsu are articulated as varying not only in volume but also in tonal quality as the kuchi-showa phonetic notation implies. According to the performer's description, Don has a more "concentrated" sound and "less overtone" than Tsu.

In analyzing musical articulation styles using NB, staccato/marcato and legato were more easily distinguished in Don rather than Tsu contexts. This suggests that there may be something inherent in the Tsu sound that is more articulation dependent than Don. IB did better than NB for attack classification. Solving a Gear4/Level5 4 class problem (Don-marcato vs. Don-legato vs. Tsu-staccato vs. Tsu-legato), IB was 97.6% accurate. The optimized parameters (after 100 iterations) are shown in Table 4. It is possible that IB outperformed the other methods for predicting stroke style because of IB's ability to memorize more stroke variations and to tune its spectrogram representation to fit the given problem.

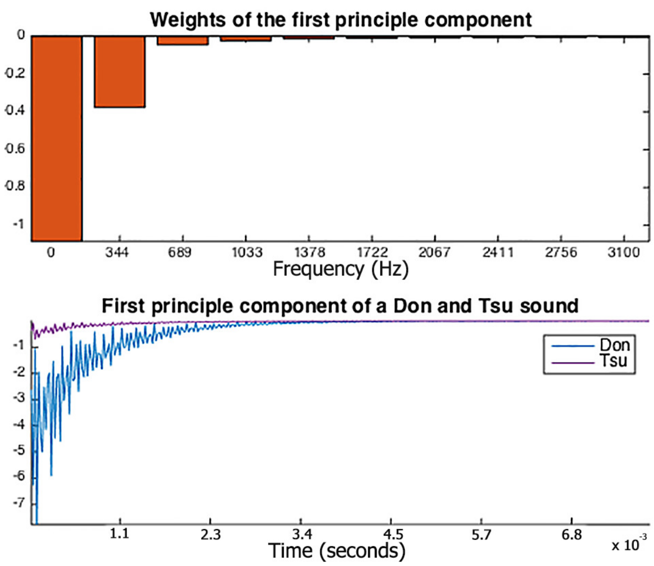

Figure 4: PCA weights and coefficients of Don and Tsu sounds

Parameter	Value
SampleTrim	0.0424s
DampingRatio	0.263
MinBandFrequency	817Hz
MaxBandFrequency	12337Hz
SpectraSampleRate	8.34
SpectraNumBands	580

Table 4: Parameters used for Don-marcato vs. Don-legato vs. Tsu-staccato vs. Tsu-legato classification for gear 4 level 5

Visual Analysis

Using HMMs, our experiments showed nearly perfect classification of gear/level combinations, Don/Tsu, and musical articulations (legato, stacatto, and marcato). The analyses showed that the velocity of movement was the key feature in distinguishing between most of the classes. However, using velocity in combination with the other movement features, such as acceleration, speed, and cross product, could also help with improving the accuracy of the models.

To analyze the performance of taiko drum-strokes more thoroughly, we examined the mean velocity of Don and Tsu strokes and their musical articulations during the attack and rebound of the stroke. Since there are only two major axes involved in performing the basic taiko drum-strokes, we analyzed the Y-axis velocity (vertical) and the Z-axis velocity (camera). In the Kinect's coordinate system, the camera is positioned at the origin, and the Z-axis represents distance from the XY plane of the camera.

Figure 5 compares the average velocity of Don and Tsu strokes in Y and Z directions. On the Z-axis we observe positive average velocity during the attack of the stroke, and negative average velocity during the rebound of the stroke. This suggests that both actions are performed by moving the arm towards the body. We can also observe that in comparison with the Tsu strokes, the Don strokes have lower velocity in the Y-axis and higher velocity in the Z-axis. We can interpret this as the drumstick in Don strokes tends to hit the drum tangent to the surface, while the Tsu strokes use more force in the vertical direction. These findings align with the taiko performer's description, describing Don as having more "bounce" than Tsu, and Tsu being more "controlled." These findings also closely relate to previous work by Dahl and Altenmüller that found correlations between stroke velocity, rebound, and the perceived timbre quality of the sound produced [12].

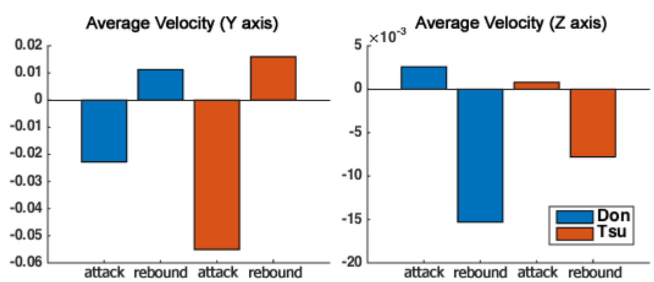

Figure 5: Comparison of the average velocity of performing Don and Tsu sounds

Figure 6 shows that the Don-marcato articulation has higher Y-velocity during the attack of the stroke compared to Don-legato, and lower Y-velocity after the stroke during the rebound. We can interpret that Don-marcato strokes require more control in the vertical position after the attack, while Don-legato strokes tend to be more free with a rounder bounce after the attack in the vertical direction. In the Z-direction, Don-legato has higher velocity compared to Don-marcato, both in the attack and rebound of the stroke. We can interpret this to mean that in Don-legato performance, the drumstick tends to hit the drum tangent to the surface, while in Don-marcato it is more perpendicular. Furthermore, Don-legato performance has a positive Z-velocity both in the attack and in the rebound of the stroke, both of which are negative in Don-marcato. A positive Z-velocity indicates movement opposite to the camera and negative velocity indicates movement towards the camera. This means that Don-legato strokes are mostly executed towards the performer, while Don-marcato strokes are mostly executed away from the performer.

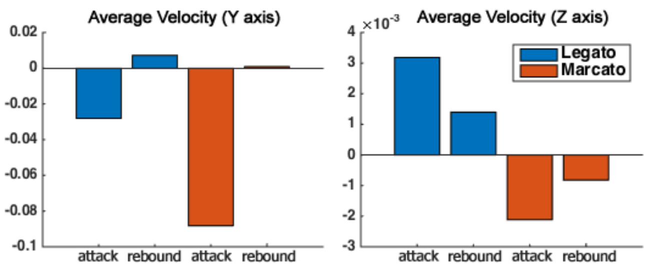

Figure 6: Comparison of the average velocity of performing legato and marcato articulations of Don

In contrast to Don strokes, Figure 7 shows that Tsu-staccato has higher velocity than Tsu-legato in both Y and Z directions. Tsu-legato has negative average Z-velocity in both the attack and rebound of the stroke, while Tsu-staccato has positive Z-velocity in the attack and negative Z-velocity when rebounding. These results demonstrate that Tsu-staccato strokes are performed by attacking the drum towards the body and bouncing back in the opposite direction (away from the body), while Tsu-legato strokes are performed primarily away from the body (for both the attack and the rebound of the stroke). These findings also correlate to the performer's experience and description of the musical articulations, describing legato as "rounder or more open" and marcato/staccato strokes as more "direct attacks that are much faster than legato and have a firmer grasp."

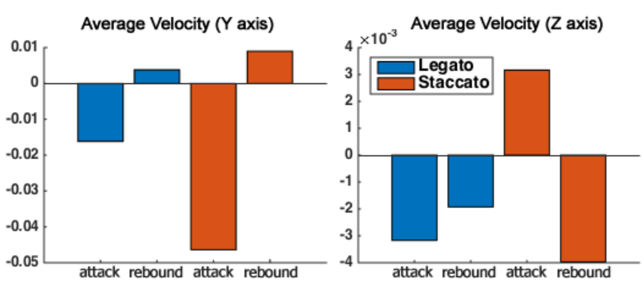

Figure 7: Comparison of the average velocity of performing legato and staccato articulations of Tsu

LIMITATIONS AND FUTURE WORK

While our results are encouraging, we have only described the first step in answering our main research questions: 1) Can visual qualities of taiko be heard in the sound and 2) Can sonic qualities of taiko be seen in the movement? In order to answer these questions we will need to enlarge our dataset and and use multi-modal analysis techniques to uncover more specific relationships between the sonic and visual data. Furthermore, we will not know if the expressive taiko features we have recognized are performer independent until we train the machine-learning models on multiple performers. Sound qualities will also vary depending on the particular drum and drumsticks used, as articulated by Dahl (2011) [11]. In future work we will continue to explore these variables and understand their influence on recognition.

We were not able to distinguish all combinations of gears and levels. This is mostly due to the exploratory nature of our data collection methods at this stage of our research. In future work we will gather a more complete set of gear and level combinations to throughly test if both gears and levels can be recognized. If the gear and level of each stroke can be detected in a sound, this can lead to the development of robust pedagogical and notating tools for taiko that can detect the stroke's position through audio data alone.

In the next stages of this research we plan to add additional labels to our taiko dataset. While dance and movement notation systems can help inform researchers on how to annotate and recognize important expressive features in movement, there is a need to understand how these expressive features can be generalized and articulated across movement styles. Just like computer programming languages, each notation system has been created for ease in recording specific styles of dance. Many of these notation systems either leave out important expressive qualities that are intrinsically tied to the technique or are too general to capture a technique's specificities. In adding more labels to our taiko dataset we can understand expressive qualities through multiple perspectives and build a bridge between taiko fundamentals and expressive qualities of movement in other contexts.

CONCLUSION

Extracting expressive features from movement data has been a challenging task in third-wave HCI design. We propose that codified artistic forms, like taiko, can offer valuable insight into how we can measure, record, and classify expressive movement qualities for automatic recognition because they are a living record of perceptually evident expressive features in human motion. The applications for this research are numerous and include the design of pedagogical systems for taiko technique and artistic explorations. Taiko, in particular, is a useful technique to explore expressive movement because sound and movement are closely integrated and equally emphasized, providing two modalities to explore expressive features. We propose in future work that annotating taiko performance through multiple perspectives can provide insight into how to generalize taiko's expressive movement features so they can be recognized and extracted in other movement contexts.

ACKNOWLEDGMENTS

This research was supported by the Social Sciences and Humanities Research Council of Canada and the Canada Council for the Arts. We would also like to thank Doug Masuhara and Jason Overy for sharing their knowledge of taiko with us and the anonymous reviewers for their helpful comments on an earlier draft of this paper.

REFERENCES

1. Aha, D., Kibler, D., and Albert, M. Instance-based learning algorithms. *Machine Learning 6*, 1 (1991), 37–66.
2. Alaoui, S. F., Caramiaux, B., Serrano, M., and Bevilacqua, F. Movement qualities as interaction modality. ACM Press (2012), 761.
3. Aronoff, J., Woike, B. A., and Hyman, L. M. Which are the stimuli in facial displays of anger and happiness? configurational bases of emotion recognition. *Journal of Personality and Social Psychology 62*, 6 (1992), 1050 – 1066.
4. Atkinson, A. P., Tunstall, M. L., and Dittrich, W. H. Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition 104*, 1 (2007), 59 – 72.
5. Bouënard, A., Wanderley, M. M., and Gibet, S. Gesture control of sound synthesis: Analysis and classification of percussion gestures. *Acta Acustica united with Acustica 96*, 4 (2010), 668–677.
6. Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., and Haggard, P. Seeing or Doing? Influence of Visual and Motor Familiarity in Action Observation. *Current Biology 16*, 22 (Nov. 2006), 2277.
7. Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., and Volpe, G. Multimodal analysis of expressive gesture in music and dance performances. In *Gesture-based communication in human-computer interaction*. Springer, 2004, 20–39.
8. Caramiaux, B., Wanderley, M. M., and Bevilacqua, F. Segmenting and parsing instrumentalists' gestures. *Journal of New Music Research 41*, 1 (2012), 13–29.
9. Cook, P. R. Physically informed sonic modeling (phism): Synthesis of percussive sounds. *Computer Music Journal* (1997), 38–49.
10. Dahl, S. Playing the accent - comparing striking velocity and timing in an ostinato rhythm performed by four drummers. *Acustica United with Acta Acustica 90*, 4 (2004), 762–776.
11. Dahl, S. Striking movements: A survey of motion analysis of percussionists. *Acoustical Science and Technology 32*, 5 (2011), 168–173.
12. Dahl, S., and Altenmüller, E. Motor control in drumming: Influence of movement pattern on contact force and sound characteristics. *Journal of the Acoustical Society of America 123*, 5 (2008), 3122.
13. Demoucron, M. *On the control of virtual violins-Physical modelling and control of bowed string instruments*. PhD thesis, Université Pierre et Marie Curie-Paris VI; Royal Institute of Technology, Stockholm, 2008.
14. Gross, M. M., Crane, E. A., and Fredrickson, B. L. Effort-Shape and kinematic assessment of bodily expression of emotion during gait. *Human Movement Science 31*, 1 (Feb. 2012), 202–221.
15. Jensenius, A. R., Wanderley, M. M., Godøy, R. I., and Leman, M. Musical gestures: Concepts and methods in research. In *Musical Gestures: Sound, Movement, and Meaning*, R. I. Godøy and M. Leman, Eds. Routledge, New York, 2010, 12–35.
16. John, G. H., and Langley, P. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc. (1995), 338–345.
17. Keefe, D. H. Physical modeling of wind instruments. *Computer Music Journal* (1992), 57–73.
18. Kleinsmith, A., Silva, P. R. D., and Bianchi-Berthouze, N. Cross-cultural differences in recognizing affect from body posture. *Interacting with Computers 18*, 6 (2006), 1371 – 1389. Special Issue: Symbiotic Performance between Humans and Intelligent Systems.
19. Levy, J., and Duke, M. Laban Movement Analysis in the study of personality, emotional state and movement style: An exploratory investigation of the veridicality of body language. *Individual Differences Research 1*, 1 (2003), 39.
20. Masurelle, A., Essid, S., and Richard, G. Multimodal classification of dance movements using body joint trajectories and step sounds. In *Image Analysis for Multimedia Interactive Services (WIAMIS), 2013 14th International Workshop on* (July 2013), 1–4.
21. Mentis, H. M., and Johansson, C. Seeing movement qualities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, ACM (New York, NY, USA, 2013), 3375–3384.
22. Overy, J. Uzume taiko. http://www.uzume.com/, 2014.
23. Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. Perceiving affect from arm movement. *Cognition 82*, 2 (2001), B51 – B61.
24. Rabiner, L. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*, 2 (1989), 257–286.
25. Shiratori, T., Nakazawa, A., and Ikeuchi, K. Detecting dance motion structure through music analysis. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference* (May 2004), 857–862.
26. Studd, K., and Cox, L. *EveryBody is a Body*. Dog Ear Publishing, 2013.
27. Teixeira, E. C., Loureiro, M. A., Wanderley, M. M., and Yehia, H. C. Motion analysis of clarinet performers. *Journal of New Music Research* (2014), 1–15.
28. Varian, H. *The way of taiko*. Stone Bridge Press, 2005.
29. Young, D., and Fujinaga, I. Aobachi: A new interface for japanese drumming. In *Proceedings of the 2004 Conference on New Interfaces for Musical Expression*, NIME '04, National University of Singapore (Singapore, Singapore, 2004), 23–26.