

THE SILENT DRUM CONTROLLER: A NEW PERCUSSIVE GESTURAL INTERFACE

Jaime Oliver and Mathew Jenkins

University of California, San Diego. Department of Music
Center for Research in Computing and the Arts, CRCA-CALIT2
joliver1@ucsd.edu // mbjenkin@ucsd.edu

ABSTRACT

This paper seeks to explain the Silent Drum Controller, designed and created for real-time computer music performance through the experiences of the authors as computer musician and percussionist. Video technology is used to extract parameters from shapes in an elastic drumhead that are tracked and mapped in the Pd/GEM environment. This system outputs raw data for further feature extraction, processing, and mapping in audiovisual work

1. INTRODUCTION

Our collaboration began with a desire to create a versatile control environment for a percussionist. The visual dimension of a percussionist's performance is rich with detail and subtlety. Through designing a controller that would utilize and build off of a percussionists pre-existing gestural vocabulary [12] we could extend the possibilities for gestural control of sound. New control paradigms would then allow us to elaborate upon the unexplored terrains within computer music and percussion performance.

The Silent Drum Controller is a transparent drum shell with an elastic head. As one presses it, the head deforms and a variety of shapes with peaks are created reflecting the shape of the mallet or hand. These shapes are captured by a video camera that sends these images to the computer, which analyzes them and outputs the tracked parameters. A diagram and picture of the drum can be seen in Figures 1 and 2.

In this paper we will outline the process for the development of the prototype. Section 2 discusses the most relevant prior work and briefly compares it to the Silent Drum Controller. Section 3 presents the aesthetic motivations that lead us to its design. Section 4 explores perceptual phenomena associated with latency as it is relevant to real-time video analysis and gesture. Section 5 explains the prototype and possible future developments.

2. PRIOR WORK

Miranda and Wanderley [11] classify controllers according to the degree of resemblance to acoustic instruments. They define instrument-inspired gestural controllers as “gestural controllers [that are] inspired by existing instruments

or that intend to overcome some intrinsic limitation of the original models. [They] do not attempt to reproduce them exactly.” Apart from bowing and several artifices, the intrinsic limitation of drums and most drum controllers is their inability to create continuous sonic events.

There are many approaches to drum controllers, such as Machover's drum-boy [5][6] and several commercial drum pads. These controllers focus almost exclusively on acquiring onsets and intensity of strike, except for the Roland V-Drum series that track radial position in the drum and the Korg Wave Drum, which allowed pressure sensing.

It is also worth mentioning Buchla's controllers Marimba Lumina, that allowed mallet identification and position of strike in the bar, and the Thunder controller, that provided position and pressure sensing[1]. Max Mathews Radio Baton [9][10] and Mathews/Boie/Schloss Radio Drum are sophisticated control environments that utilize a gestural language similar to that of percussion. One uses two radio batons tracked through capacitive sensing to strike a pad. This controller provides data for the (x, y, z) coordinates of the mallets in space and the (x, y) positions in the pad. Therefore, the basic features of traditional percussive gestures can be measured within this control environment. Another controller that uses a malleable surface was created by [14]. It also allows to track one hand and fingers. A thorough review of percussion controllers can be found in [1].

Our instrument-inspired controller provides (x, z) coordinates when the head is struck or pressed. The elastic head allows us to create shapes through its deformation. Complex shape acquisition and the possibility of extracting several discrete events are some of its advantages. This is in stark contrast to an acoustic drum and most controllers. Within this control environment one can manipulate continuous sounds through a new gestural vocabulary.

3. AESTHETIC MOTIVATIONS

The design of the Silent Drum Controller was motivated by various aesthetic considerations. It was designed primarily with the possibility of dissociating gesture with sound. Furthermore, we wanted to advance the pre-existing state of drum controllers into a richer control environment.

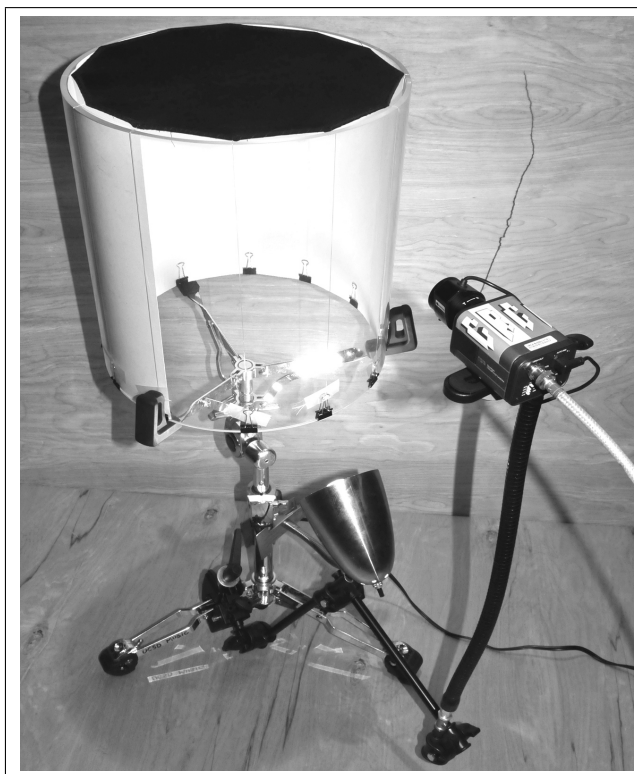


Figure 1. This is the prototype of the Silent Drum.

This environment allows independent control of up to 22 variables and its derivations. Feature extraction can obtain multiple discrete events even while holding a continuous gesture.

Traditional acoustic instruments generally have a direct correspondence between gesture and sound that we can anticipate. The Silent Drum controller destabilizes this vision of instrumental performance. The malleability of the head allows not only for percussion gestures, but for hand and finger tracking. In this way, the controller acts as the limits, but also as an extension of the human body.

In contrast to most hard material pads, an audible sound is not emitted from the Silent Drum Controller when struck. This enables us to capture gestures and either utilize them immediately sonically or store them for future sound transformations. In real-time computer music, sound transformation is frequently automated through cue lists triggered by score followers. It can be difficult to build expectations within this listening environment. Our approach uses discrete events from the drum to drive the score and produce changes in mappings. The Silent Drum Controller can visually inform the listener of how the gesture is influencing the sonic outcome.

Lastly, most drum controllers are extremely limited. With few exceptions, commercial drum controllers only detect onsets and measure intensity of strike or ‘velocity’. This reduces the space for subtle, virtuosic control and no possibility of controlling continuous sound transformations. Mathews’ controllers have a wide palette of variables to control. However, one can’t perform his controllers with percussion mallets. This leaves little space

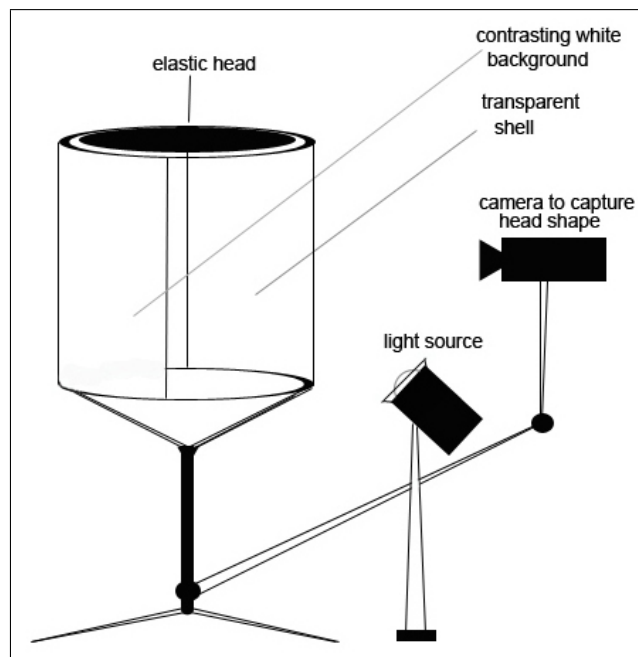


Figure 2. The basic elements of the Silent Drum.

for integration within more traditional percussion set-ups.

4. PERCEPTUAL CONSIDERATIONS

Mapping and synchronicity are implied features of acoustic instruments. Sensing and processing an environment is inherently latent and asynchronous. The use of video based controllers in live computer music requires the consideration of cross modal perception of synchronicity. The perception of synchronicity between a gesture and its sound requires events to happen within certain perceptual boundaries across modes of perception. Haptic, auditory, and visual channels are usually engaged during a performers interaction with a controller. The audience perceives the performance through the auditory and visual channels. Two types of perceptual phenomena, discrete and continuous events, characterize both of these interactions.

The audience receives visual and sonic stimuli from the performance. Visual stimuli are received almost immediately. In most performance settings sound arrives with a delay of approximately 3 milliseconds/meter. The use of a computer increases this delay. According to [4] the tolerable delay for sonic and visual discrete events to be perceived as synchronous by an observer is 45 milliseconds.

The performer receives haptic and sonic feedback from the controller. [4] reported a tolerable delay of 42ms between haptic and sonic feedback for discrete events. In the case of continuous events without tactile feedback, [7] calculated the tolerable delay between a gesture and a sound at 30ms for a sinusoid and as high as 60ms for a sinusoid with vibrato. We agree with [8] that latency toleration varies with the mapping used and other variables.

[2] calculated that “in piano performance the delay between pressing the key and the onset of the note is about

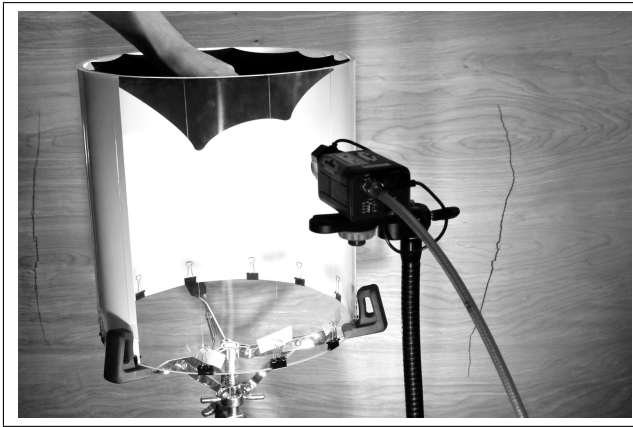


Figure 3. The deformed drumhead being pressed with a hand. The shape is captured by the camera. A lighting system provides enough contrast for fast tracking.

100 ms for quiet notes and 30 ms for forte notes.” Audience and performers develop anticipatory mechanisms to compensate for latency. “By this we mean that the arrival of sensory input through one sensory channel causes a neural module to prepare for (that is, to predict) the arrival of related information through a different sensory channel” [4]. As we will explain later, a low and consistent latency can be achieved, which could provide the performer with the adequate conditions to develop an anticipation strategy.

5. PROTOTYPE

We wanted our prototype to look like a drum to which we could attach an elastic drumhead and a camera. This wasn't strictly a visual requirement. Drums have a developed infrastructure that we could take advantage of, such as the availability of stands, drumhead rings and shells, so we wouldn't need to fabricate new equipment. It also provides flexibility within percussion setups.

In order to adapt to the needs of video tracking, we needed a stable, transparent shell and a white reflective background that could allow for controlled lighting. We required the material for the head to be elastic, to have a contrasting dark color, and to resist deformation and breaking. We are currently using spandex. Figure 3 shows an image of the drum being used.

We are using a video camera with a region of interest (ROI) of 620x260 pixels. This ROI gives a spatial resolution of twice MIDI for the vertical axis and almost 5 times MIDI for the horizontal axis. To obtain an adequate temporal resolution, a 200fps uncompressed firewire camera is used. This gives a latency variation or jitter of 5 ms. This capture frequency is still limited in capturing percussive gestures such as flams as stated by [16].

Currently, the video and audio processes are done in one computer. Pd's audio latency could take as low as 10 ms. This gives us a total system latency of 12.5 ± 2.5 milliseconds, well within the perceptual latency tolerance

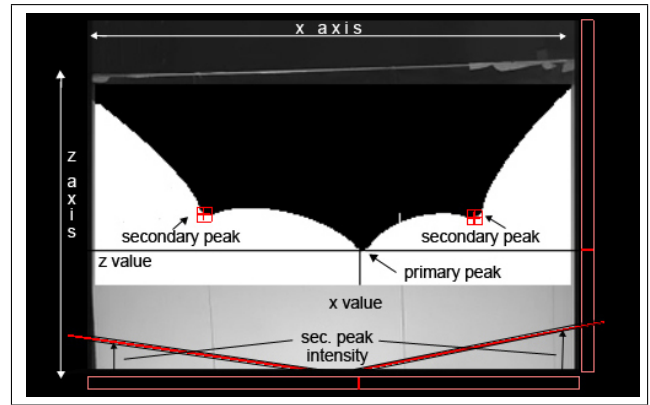


Figure 4. Figure 4. The output image of pix_drum

for synchrony found by [4] and [8].

The video tracking algorithm was implemented in the Pd/GEM environment [3][13] and has been called pix_drum. Figure 4 shows the output image of pix_drum.

The algorithm works roughly in the following way:

- Step 1 Threshold the image so that the drumhead becomes black and the background white. This is a robust mechanism given the controlled lighting.
- Step 2 Record a histogram of the vertical values; z coordinate indexed by the x coordinate (number of black pixels per column).
- Step 3 Determine the location of the primary peak, in terms of the x coordinate (the index of the highest value in the histogram) and the z value (the highest value itself).
- Step 4 Determine the total black area (the total number of black pixels) and the black area on each side of the primary peak. This is a good measure of the intensity of secondary peaks when compared to what the area would be with only the primary peak.
- Step 5 Determine the position (x, z) of the secondary peaks. Starting from each side of the histogram towards the primary peak, we test for contiguous decreases in the histogram as a convexity test. (If there were no secondary peaks, the histogram should show a continuous increase until it reaches the primary peak).

The basic output of pix_drum is 4 raw streams of data. This includes the (x, z) positions of the primary peaks and the area covered on each side of the central peaks. Up to 18 secondary peaks are outputted in order of appearance. Filtering, interpolation and feature extraction can be made in a patch outside the object or in a separate computer altogether. This is beneficial, because it reduces the amount of data to be transmitted in addition to giving the end user the option to extract their own features [17]. We've developed separate patches to filter, interpolate the data, and extract discrete features such as onset,

offset, velocity, and inflection points. Output images are optional so that cpu time can be saved, however we find that they are helpful for debugging and as optional visual feedback. The code and patches for pix_drum, as well as videos with specific sound mappings, can be accessed at <http://crca.ucsd.edu/silent>. Current feature extraction provides discrete events other than onsets such as direction changes. Discrete events can be obtained while controlling continuous events.

In the future, we will include a second camera on the controller to obtain depth (y axis, not depth of strike) using stereo correlation. A complimentary and ongoing project provides the position of mallets in space through tracking the colors of the mallet heads. This could also provide significant information to anticipate an onset in the drum [12] and reduce latency in it's detection.

6. CONCLUSIONS

The continuous improvement and cost in hardware has made it possible to use video as a sensor for live computer music controllers. This provides us with flexible tracking strategies and this trend can only improve in the future. The live computer music community could benefit from determining latency toleration boundaries in variable contexts and mappings, as well as from determining the effects of latency variation on the development of anticipatory mechanisms. The Silent Drum Controller opens up new possibilities for gestural control of sound, through the acquisition of traditional and non-traditional percussive gestures. It also provides an alternative to drum pads and the possibility of using percussion mallets and hands. This controller opens doors to new aesthetical explorations for percussion and live electronics.

7. REFERENCES

- [1] Aimi R.M. "New Expressive Percussion Instruments" *PhD Thesis Massachusetts Institute of Technology* 2002
- [2] Askenfelt, A. and Jansson, E.V. "From touch to string vibrations. I: Timing in the grand piano action" *The Journal of the Acoustical Society of America* 1990
- [3] Danks, M., Real-time image and video processing in GEM *Proceedings of the International Computer Music Conference Thessaloniki, GREECE*, 1997.
- [4] Levitin, D. J., Mathews, M. V., and MacLean, K., The Perception of Cross-Modal Simultaneity.. *Proc. of International Journal of Computing Anticipatory Systems* Belgium, 1999.
- [5] Machover, Tod, Classic Hyperinstruments: A Composer's Approach to the Evolution of Intelligent Musical Instruments. <http://brainop.media.mit.edu/Archive/Hyperinstruments/classichyper.html> 1997
- [6] Machover, Tod, Hyperinstruments <http://web.media.mit.edu/~tod/Tod/hyper.html> 1998.
- [7] Mäki-Patola, T., Hämäläinen, P. Latency Tolerance for Gesture Controlled Continuous Sound Instrument without Tactile Feedback *Proceedings of the International Computer Music Conference* Miami, USA, 2004.
- [8] Mäki-Patola, T. Musical Effects of Instrument Latency *Proceedings of the Suomen Musikututkijuiden 9. valtakunnallinen symposium* Jyväskylä, Finland, 2005.
- [9] Mathews, M. V. "The Conductor Program and Mechanical Baton," In M. V. Mathews, ed. *Current Directions in Computer Music Research* MIT Press. Cambridge, Massachusetts, 1989.
- [10] Mathews, M. V. "The Radio Drum as a Synthesizer Controller." *Proceedings of the International Computer Music Conference* San Francisco, USA, 1989
- [11] Miranda, E.R., Wanderley, M. *New digital musical instruments: control and interaction beyond the keyboard*. A-R Editions, Middleton, Wis., 2006
- [12] Puckette, M., Settel, Z. Nonobvious roles for electronics in performance enhancement *Proceedings of the International Computer Music Conference* San Francisco, USA, 1993.
- [13] Puckette, M. "Pure Data" *Proceedings of the International Computer Music Conference* San Francisco, USA, 1996.
- [14] Vogt, F. and Chen, T. and Hoskinson, R. and Fels, S. "A malleable surface touch interface" *International Conference on Computer Graphics and Interactive Techniques* ACM Press New York, NY, USA, 2004
- [15] Wessel, D., Wright, M. Problems and Prospects for Intimate Musical Control of Computers *Proceedings of the Conference on New Interfaces for Musical Expression, NIME* Seattle, USA, 2001.
- [16] Wright, M. Problems and prospects for intimate and satisfying sensor-based control of computer sound *Proceedings of the Symposium on Sensing and Input for Media-Centric Systems, SIMS* Santa Barbara, USA, 2002.
- [17] Zicarelli, David. Communicating with Meaningless Numbers *Computer Music Journal* 15(4): 74-77 1991