SQUISHBOI: A Multidimensional Controller for Complex Musical Interactions using Machine Learning

Odie DeSmith
California Institute of the Arts
24700 McBean Parkway
Valencia, CA 91355
marceldesmith@
alum.calarts.edu

Andrew Piepenbrink
California Institute of the Arts
24700 McBean Parkway
Valencia, California 91355
andrewpiepenbrink@
alum.calarts.edu

Ajay Kapur
California Institute of the Arts
24700 McBean Parkway
Valencia, California 91355
akapur@calarts.edu

ABSTRACT

We present SQUISHBOI, a continuous touch controller for interacting with complex musical systems. An elastic rubber membrane forms the playing surface of the instrument, while machine learning is used for dimensionality reduction and gesture recognition. The membrane is stretched over a hollow shell which permits considerable depth excursion, with an array of distance sensors tracking the surface displacement from underneath. The inherent dynamics of the membrane lead to cross-coupling between nearby sensors, however we do not see this as a flaw or limitation. Instead we find this coupling gives structure to the playing techniques and mapping schemes chosen by the user. The instrument is best utilized as a tool for actively designing abstraction and forming a relative control structure within a given system, one which allows for intuitive gestural control beyond what can be accomplished with conventional musical controllers.

Author Keywords

tactile interaction, gesture, mapping, machine learning

CCS Concepts

•**Human-centered computing** → **Gestural input;**
•**Applied computing** → *Sound and music computing;*
•**Hardware** → *Sensor devices and platforms;*

1. INTRODUCTION

Modern communication protocols such as MIDI and OSC have allowed for the development of parametrically driven hardware performance interfaces for controlling large-scale multichannel systems in real time. These systems have led to a blending of traditional instrumental performance techniques with multitrack mixing techniques common in DJing. However, we believe that for this blending to better realize the potential of modern computer processors, we will need better interfaces — interfaces with the ability to link simple gestural interactions with high-dimensional control data. SQUISHBOI's array of time-of-flight (TOF) distance sensors and its rubberized touch surface offer the potential for such control within a single tactile interface. This controller takes aim at the space to the left of the traditional keyboard, where the mod wheel and pitch bend have sat more or less unchanged for decades; even more sophisticated interfaces like the left-hand 'drawer' of the Ondes Martenot [5] and its modern descendant, the Expressive E Touché[1], offer only a few axes of expression. SQUISHBOI provides its user with a unique gestural interface designed to inspire new sounds and musical interactions within complex parametric systems.

2. BACKGROUND

Figure 1: Installation at Supplyframe DesignLab.

2.1 Motivation

Many hardware interfaces are at a stage where they are struggling to keep up with the flexibility and modulation potential of computer music software. The advent of sophisticated controllers like the Ableton Push demonstrates that in order to control the high level of complexity in modern DAWs, controllers must become systems in and of themselves. This new paradigm in computer music interaction carries on the tradition of hardware production platforms such as the Akai MPC; such tools made it possible for individuals to perform complex musical arrangements with a user-programmable method of interaction. While these systems have liberated individuals to both create and perform musical compositions in new and exciting ways, they also often fail to demonstrate concrete gestural relevance to the

---

[1]https://www.expressivee.com/buy-touche

sonic output. When there is no directly visible method of interaction with a musical system, audience members are forced to speculate as to how sounds are being produced [6]. What if we could engage with smart controllers without sacrificing performativity? What if there were a controller that could highlight the importance of physicality and muscle memory and also offer high-level control? These were the questions that led us to design and implement a gesturally driven controller for real-time interaction with complex musical systems.

2.2 Prior Art

SQUISHBOI stands on the shoulders of giants in the history of multitouch interfaces for intimate realtime control of music. Since the advent of the Monome[2] in 2006, grid-based MIDI controllers have become ubiquitous, affordable, and familiar. Other instruments forgo an array of discrete buttons in favor of a more versatile single surface, including the Sensel Morph[3], Madrona Labs Soundplane [3], and the Reactable [4]. Most of these instruments also possess a third (Z) axis in the form of pressure sensitivity; what sets our instrument apart is that its typical depth excursions are very large, on the order of the size of the performer's hand. Despite their differences, the concept of controlling large-scale computer music systems from one tactile playing surface remains a consistent and common theme among all these instruments. MIDI Polyphonic Expression (MPE) controllers attempt to add additional layers of control to familiar interfaces, most notably the conventional piano keyboard. SQUISHBOI takes a different approach, embracing the unusual cross-couplings inherent in its elastic playing surface. Though these couplings make some familiar control paradigms challenging, like all instruments its couplings and constraints contribute to its character and that of the music made on it.
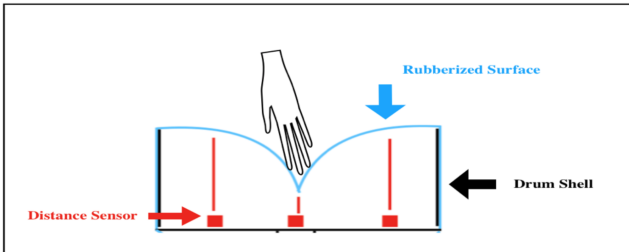
3. SYSTEM DESIGN

Figure 2: Instrument Layout.

The rubberized membrane surface is a key component of SQUISHBOI's design. We needed a material that could be easily stretched, but also one thick enough for the distance sensors to properly register measurements. Furthermore, the surface needed to be flexible yet durable, and capable of sufficient depth excursion to support the intended gestures. Cross-coupling between individual distance sensors is achieved through the inherent physical properties of the rubberized surface, which deforms over areas much wider than a touch itself. When properly tensioned for depth response the membrane is very loose compared to a normal drumhead, and makes almost no acoustic sound when played. We eventually settled on a form of industrial-grade latex sheet with the right properties for our first prototypes, which were robust enough to last through hours of testing

[2]https://monome.org/
[3]https://sensel.com/pages/the-sensel-morph

and practicing, a number of performances, and two public presentations over the course of several months[4].

The physical enclosure for the controller was adapted from a 14"x4" snare drum. While in the future we may look to fabricate a custom enclosure for the controller, the existing form factor and hardware made the modified snare drum an appropriate choice for the initial prototype. Although this was certainly the path of least resistance, we learned quickly that the form factor of the instrument often shaped how users would interact with it. In public exhibitions of the instrument, a notable percentage of participants seemed to assume that its resemblance to a drum meant it was to be struck rather than stretched, leading to dissatisfying results. We viewed this as a strong indication that, in the future, we would need to either discourage such assumptions by redesigning the enclosure to be less drum-like, or welcome them by modifying the circuit and firmware to better respond to percussive gesture.
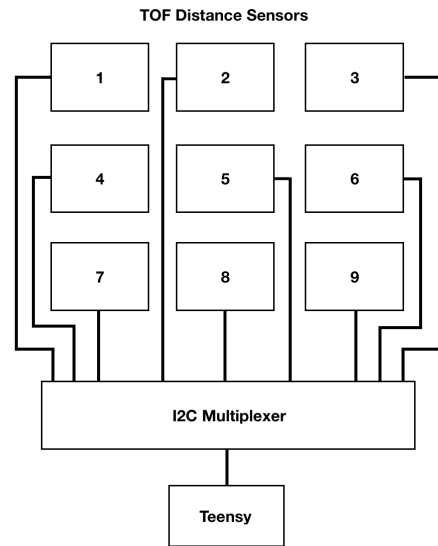
3.1 Hardware

Figure 3: Sensor System Architecture.

SQUISHBOI's functionality is achieved through a sensor array comprised of nine VL6180X[5] time-of-flight distance sensors. These sensors are used to make parallel measurements from the base of the instrument to the rubberized latex playing surface mounted above. The sensors are connected via the I2C bus to a microcontroller, where the data is converted into MIDI CC messages accessible via USB. The VL6180X is a time-of-flight distance sensor, which carries significant advantages over older distance sensing technologies. Ultrasonic sensors were off the table immediately, since the cone of sensing was too wide and would result in nonstop crosstalk between the sensors. Traditional infrared sensors were also ruled out, since their readings would be based on the amount of light returned back to the sensors' receivers, which would cause many issues with linearity and double imaging. The VL6180X has neither of these limitations, using a cutting-edge form of Micro-LIDAR that is currently being aggressively developed for use in self-driving

[4]https://vimeo.com/338113310
[5]https://www.st.com/resource/en/datasheet/vl6180x.pdf

cars, drones, and robotics. We can expect to see more musical and art-based applications of LIDAR technologies in the near future as these sensors become more ubiquitous and affordable.
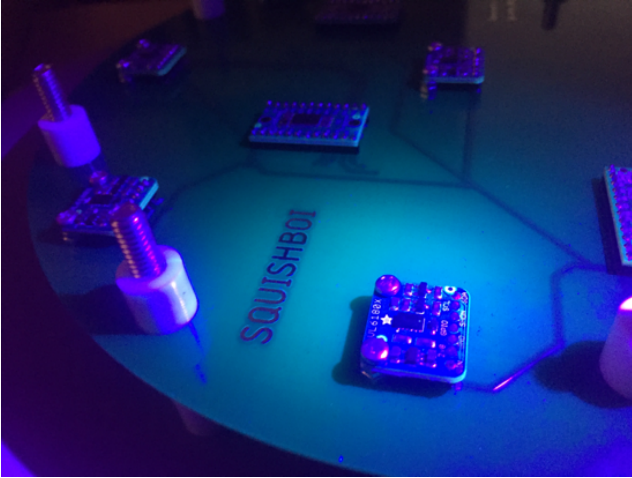

Figure 4: Circuit board with sensors.

SQUISHBOI's printed circuit board has been fabricated both to position and interface with its nine distance sensors. The rubberized membrane cover helps keep lighting conditions consistent, and also ensures adjacency between the beams of each sensor's laser, which is crucial to the intended functionality of the interface. Placing the entire circuit on one large disc-shaped board greatly improved the placement accuracy of each sensor within the array, and allowed for the shortest turnaround time between prototyping and final assembly.

3.2 Software

SQUISHBOI's multichannel sensor array and tactile rubberized membrane make it an appealing physical interface, but it is difficult to use this data directly. This makes it a perfect target for machine learning (ML) software such as Wekinator [1], which can not only make unwieldy data easier to work with, but offer entirely new behaviors like gesture recognition and abstract mappings. These types of non-linear system behaviors enable performers to forget about the governing parameters of a sonic system and focus on how their movements produce different sonic outputs. This makes apparent the intimate feedback network existing between musician and instrument, further highlighting the role of muscle memory in musical performance.

In computer music performance, a performer may need to make complex parametric changes throughout their entire system to achieve a sonic goal, often in near-realtime. The audience sees them frantically repatch a cable, stare intently at a screen, anxiously turn a knob, then pause... a flurry of activity to be sure, but collectively bearing little resemblance to what most would recognize as a musical *gesture*. The intuitive link between gesture and sound, inherent in so many familiar forms of musical performance, is rendered opaque or even nonexistent. Many computer music artists must rely on the projection of synchronized visuals to help mitigate the loss of this important multimodal relationship. Visual substitutions, however, do little to benefit the performer in terms of spatial and tactile connection, traits that are essential to most all forms of traditional instrumental performance. In [8], Wessel and Wright speak to the importance of metaphors for musical control when dealing with systems of high dimensionality. In SQUISHBOI,

we use ML to prototype and explore novel metaphors which are musically effective *and* performative, hopefully regaining some of what is so often lost when music is mediated by technology

Figure 5: Instrument Shell.

A user can easily set up SQUISHBOI for continuous gestural morphing, where output parameters are interpreted via regression in Wekinator. Its emphasis on on-the-fly learning is powerful because it means we can work very quickly, testing and exploring systems directly on the instrument with no need for external programming or development environments. What if, rather than sending a continuous distance measurement between the sensor and rubberized surface, the system output an envelope or LFO once a certain threshold was met? We could easily retrain the model and test this idea in minutes; being able to quickly answer questions about the instrument on the instrument was a guiding concept in the design of its workflow.

4. MAPPING AND INTERACTION
4.1 Integration with Other Systems

SQUISHBOI's high level of dimensionality and tactile responsiveness make it an apt interface for controlling often unwieldy systems. One such system, growing in popularity over the last decade, is the Eurorack modular synthesizer format. In conjunction with a DC-coupled audio interface, SQUISHBOI can send its relative sensor data as continuous control voltage (CV) signals for interfacing with analog systems. The relatively open format embraced by most modular architectures offers a unique entry point for cross-referenced multichannel control data. This data can be used to develop a control surface for real-time interaction with hardware synthesizer patches, or even to facilitate tactile control over the governing aspects of signal flow, such as the internal routing of modulation signals.

Even modest modular synthesizer patches can grow difficult to manage with only two hands, especially when many values are accessed only by individual buttons and knobs; X/Y joysticks and pad controllers may be more comfortable to use, but are no better at taming dozens of parameters. When such a system becomes large enough, one-to-one mappings are destined to fail no matter the interface — in order to move forward, we must adopt a strategy allowing abstracted, parallel control of many sound parameters from far fewer input streams. Without such structures in place, a user simply does not have the physical or mental bandwidth to steer the ship without feeling like a cartoonesque

one-man band. With machine learning we can add a layer which will help to manage or even expand this user's bandwidth. Our hapless performer is no longer so hapless, and may even be able to add a few tunes to their repertoire.

In SQUISHBOI such an ML layer can be made to dictate all system behavior, or to complement and work in parallel with raw distance sensor signals being used elsewhere in the patch, as illustrated in Figure 6. If an additional neural network were used to further abstract the nine TOF sensors' signals relative to a different set of user-specified training data, the output signals from this secondary network could be sent to any number of additional locations if deemed necessary. This could even include things such as generative gestural accompaniment as demonstrated in prior work like *GestureRNN* [2]. The limiting factor when it comes to the breadth of control is not the interface itself, but rather the computer processor and imagination of its user.
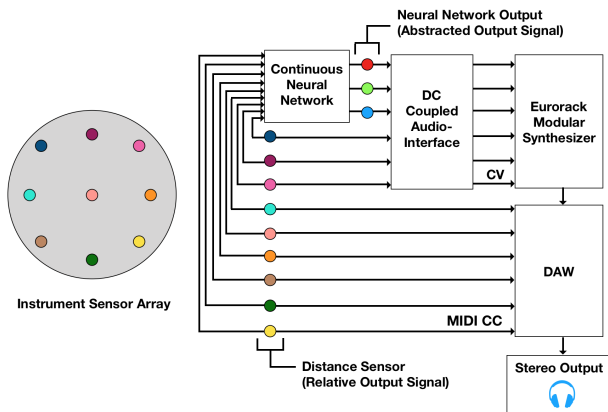

Figure 6: System Overview

4.2 Performance Techniques

SQUISHBOI enables powerful interactivity with complex sound synthesis, but its effectiveness is not just a means to an end goal of *total* or *correct* control. Indeed, in many cases its flexibility can help generate new ideas — with each new mapping, SQUISHBOI can suggest new avenues of interaction just as much as new sounds. SQUISHBOI's inherent dynamics allow for cross modulations to occur naturally, much like the natural relationship between an acoustic instrument's note-range and timbre. These dynamics can also be designed to diverge from the natural, and we have found that it really excels with complex FM synthesis in this regard. When modulating the amplitude and fine frequency controls on Ableton's Operator, we were able to produce complex and highly interesting drone textures that morphed harmonically across the different clusters of the sensor array. We also achieved startling sonic results while jamming on a mapping that we had trained to control the parameters of an arpeggiator, using real-time classification via dynamic time warping. SQUISHBOI can be thought of as a "composed instrument" — a term borrowed from David Wessel's description of *SLABS* [7], a controller similarly designed for expressive control across a user-programmable array. The main difference with SQUISHBOI lies in the tactile quality and inherent coupling of distance readings within the array that result from its flexible rubberized surface.

5. CONCLUSION

We have demonstrated a continuous controller for interacting with complex musical systems. This controller builds upon the tradition of array-based instruments, while providing users with a tactile interface built from the ground up to prioritize complex interactions along its Z axis. We have showcased how the sensor circuit network of our interface makes it a prime candidate for exploration of musical interaction with continuous neural networks. Furthermore, this work highlights how physical controllers with a high level of dimensionality can be used to reintroduce muscle memory and gestural interaction as focal points within a given computer music system.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] R. Fiebrink, D. Trueman, and P. Cook. A meta-instrument for interactive, on-the-fly machine learning. In *Proc. NIME*, pages 280–285, 2009.

[2] L. Hantrakul and Z. Kondak. Gesturernn: A neural gesture system for the roli lightpad block. In *Proc. NIME*, pages 132–137, 2018.

[3] R. Jones, P. Driessen, A. Schloss, and G. Tzanetakis. A force-sensitive surface for intimate control. In *Proc. NIME*, pages 236–241, 2009.

[4] S. Jordà, M. Kaltenbrunner, G. Geiger, and R. Bencina. The reactable. In *In Proceedings of the International Computer Music Conference (ICMC 2005*, pages 579–582, 2005.

[5] J. Loriod. *Technique de l'onde electronique type Martenot*, volume 1: Le Clavier. Alphonse Leduc Editions Musicales, Paris, 1987.

[6] W. A. Schloss. Using contemporary technology in live performance: The dilemma of the performer. *Journal of New Music Research*, 32(3):239–242, 2003.

[7] D. Wessel, R. Avizienis, A. Freed, and M. Wright. A force sensitive multi-touch array supporting multiple 2-d musical control structures. In *Proc. NIME*, pages 41–45, 2007.

[8] D. Wessel and M. Wright. Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 26(3):11–22, 2002.