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

Odie De Smith  
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, CA 91355  
andrewpiepenbrink@alum.calarts.edu

Ajay Kapur  
California Institute of the Arts  
24700 McBean Parkway  
Valencia, CA 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 and its modern descendant, the Expressive E Touche, 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

### 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

---

1https://www.expressiveee.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 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, Madrona Labs Soundplane, and the Reactable. 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

![Image](https://example.com/image.png)

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

The physical enclosure for the controller was adapted from a 14”x14” 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

![Image](https://example.com/image.png)

**Figure 3: Sensor System Architecture.**

SQUISHBOI’s functionality is achieved through a sensor array comprised of nine VL6180X 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 cars.
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.

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.

### 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.

### 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.

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
The authors would like to thank Marijke Jorritsma and Christine Meinders, whose collaborative efforts have been a guiding force in both the aesthetic and technical development of this project. We would also like to thank the CalArts community as a whole for fostering a welcoming and supportive network wherein this type of work can thrive.

7. REFERENCES