SonoCube: A Handheld Motion-Responsive Sound Object

Szymon Walendowski
swalendowski@berklee.edu
Berklee College of Music
Electronic Production and Design
Boston, MA, USA

Akito van Troyer
avantroyer@berklee.edu
Berklee College of Music
Electronic Production and Design
Boston, MA, USA

Abstract

SonoCube is a self-contained handheld sound object that produces generative ambient music controlled through physical interaction. The 70mm cube responds to touch, motion, and orientation: capacitive touch triggers melodic patterns, while shaking, tilting, and rotating modulate delay, reverb, and filter parameters. A dual-microcontroller architecture separates sensing from audio synthesis, with an Adafruit QT Py handling IMU data processing and a Teensy performing DSP. The device runs on battery power with an integrated speaker, requiring no external equipment. This paper presents the design rationale, technical implementation, and preliminary observations from user interactions, exploring how ambient synthesis can support continuous gestural interaction within a minimal handheld form. The case study examines how reducing interface elements to motion and touch interacts with ambient synthesis, and reports preliminary observations on user interaction.

Keywords

handheld instrument, motion control, capacitive touch, embedded audio

1 Introduction and Related Work

Handheld devices and sound objects can offer intimate physical engagement without the complexity of traditional instruments [3, 6, 11, 13]. The cube form factor has particular appeal: it is graspable, orientable, and immediately suggests interaction through manipulation. Cube-based interfaces have appeared in both collaborative sound design [8] and emotion-focused musical control [5]. More broadly, the development of standalone embedded instruments with integrated synthesis has been explored as a way to create more immediate, self-contained musical experiences [10].

The project emerged from a desire to create smooth, intuitive interactions between physical gesture and sound. Rather than positioning itself as an instrument for virtuosic performance or a toy for casual amusement, SonoCube exists as a study in embodied sound design, an exercise in mapping bodily movement to sonic parameters within severe physical constraints.

Design influences include Teenage Engineering's approach to playful yet refined hardware, and Kenya Hara's minimalist design philosophy [2]. The goal was not feature density but interaction quality: a small object that feels satisfying to manipulate and rewards exploration with evolving sonic textures.

SonoCube extends this territory by combining embedded synthesis with a motion-responsive, handheld form. The exploration aims to reduce interface elements to motion and touch, which combined with ambient synthesis that embraces gestural ambiguity, could support exploratory and satisfying sonic interaction without prior instruction or explanation. We position SonoCube as a design case study in minimal embodied sound interaction. The intended audience is casual users and installation contexts rather than trained performers, where immediate engagement matters more than virtuosic control.

NIME '26, June 23–26, 2026, London, UK

2 Design and Implementation

Figure 1: SonoCube held in hand, showing the 70mm translucent PETG enclosure and touch sensor LED indicator.

2.1 Physical Form

SonoCube is a 70mm cube housed in an FDM 3D-printed enclosure in PETG translucent filament[1] (Figure 1). This size allows a comfortable single-handed grip while accommodating the necessary electronics. The cube form was chosen for several reasons: it has no obvious "correct" orientation, encouraging rotation and reorientation; all faces are equivalent in shape, simplifying the mapping between physical manipulation and sonic response. At 70mm, it falls within comfortable single-hand grip dimensions, intended to support sustained interaction.

One face contains an embedded capacitive touch sensor, barely visible from the exterior. This design choice prioritizes visual minimalism but, as discussed below, creates discoverability challenges.

2.2 Technical Implementation

2.2.1 Hardware Architecture. The system uses a dual-microcontroller architecture that separates concerns between sensing and synthesis:

[1] Demo video available at https://youtu.be/NoPLpANU6IY

Adafruit QT Py RP2040 handles all sensing—IMU (gyroscope and accelerometer), capacitive touch detection—and Bluetooth Low Energy communication. The QT Py performs gravity compensation, shake detection, and face detection, outputting cleaned sensor data as JSON over a serial connection.

Teensy 4.0 with Audio Adapter Shield receives parsed sensor data and performs all synthesis using the Teensy Audio Library.

Additional components include a small D-class amplifier, 40mm speaker, and a 3.7V 350mAh lithium polymer battery. Based on typical component current draw, the estimated runtime is approximately 45–75 minutes of continuous use, varying with speaker volume. The battery is removable and charges externally via a dedicated charger in under one hour. The entire system fits within the 70mm cube enclosure with no external connections necessary during use.

2.2.2 Software and Signal Flow. Both microcontrollers run Arduino (C++) firmware. The data flow, illustrated in Figure 2, proceeds as follows:

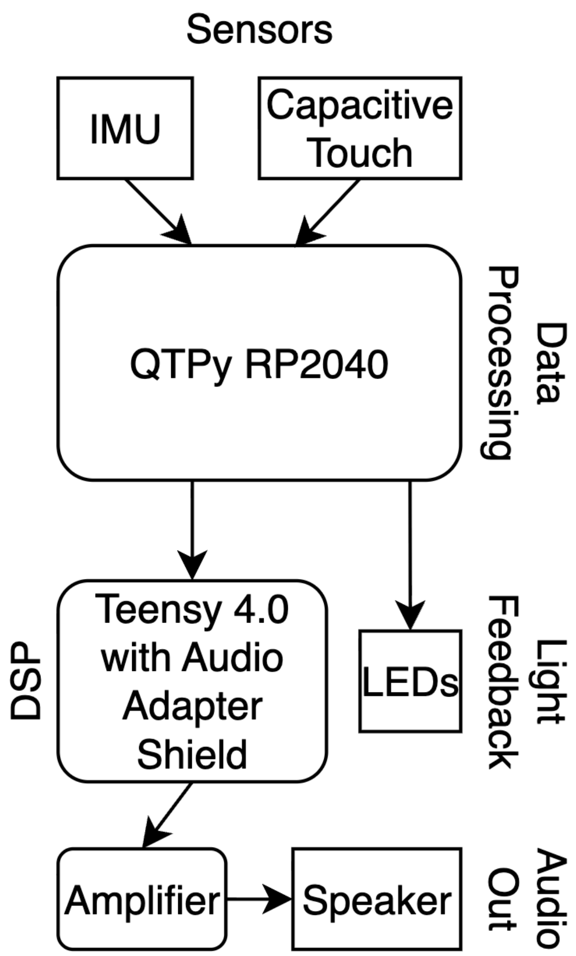

Figure 2: Signal Flow diagram.

The QT Py samples the LSM6DSO32 IMU and touch sensor at 104Hz, filters the accelerometer data to remove the gravity component, and detects shake events through jerk-based analysis. Shake detection triggers when the rate of change of acceleration exceeds 10 m/s^3, with a 500ms cooldown between events to prevent repeated triggering from a single gesture. This threshold was determined empirically during development, balancing sensitivity against false triggers from normal handling. The QT Py then packages the sensor state as JSON and transmits it over serial to the Teensy at 115,200 baud.

Table 1: Gesture-to-Sound Mapping

Gesture	Parameter	Range	Rationale
Face (orientation)	Melody note	6 pentatonic notes	Discrete faces map to discrete pitches.
Angular velocity	Reverb mix, Delay mix	0–0.8, 0–0.5	Rotation 'blurs' the sound.
Linear acceleration	Delay mix, Lowpass filter	0–0.5, 10k–20k Hz	Rapid movement brightens and blurs.
Touch	Melody trigger	on/off	Enables intentional melody.
Shake (jerk event)	Chord change	3 chords cycling	Requires intentional gesture.

The Teensy parses the incoming data and routes values to synthesis parameters. The Teensy Audio Library generates and processes 16-bit, 44.1 kHz audio in real-time, outputting to a class D amplifier driving a 4 Ω, 5 W speaker.

A secondary BLE data stream allows real-time visualization of sensor values through a simple web application (JavaScript/HTML). This proved essential during development for tuning thresholds and verifying sensor behavior, and remains available for future development work, namely a companion app running on the user's computer, displaying visuals reactive to the device's movement.

2.3 Interaction Model

SonoCube produces sound immediately upon power-on—there is no silent waiting state. The interaction model comprises three primary modalities, summarized with their sonic mappings in Table 1.

Touch. Contacting the capacitive surface triggers melodic patterns. The touch interaction is binary (touched or not touched) rather than pressure-sensitive, keeping the gestural vocabulary simple. When the sensor detects touch, a small LED lights up above the sensor's face.

Motion. The IMU (inertial measurement unit) captures acceleration and rotation. Shake gestures trigger chord changes and modulate effect intensity, while sustained tilting and rotation continuously adjust delay time, reverb mix, and filter cutoff.

Orientation. The cube's orientation determines which note of a minor pentatonic scale is available for the touch-triggered melody. A NeoPixel LED on the QT Py RP2040 shifts color based on which face points upward, six distinct colors for six faces, providing visual confirmation of orientation state.

The motion-to-sound pipeline incurs approximately 16 ms of latency for shake detection (4.8 ms average IMU sample wait, 7.8

ms serial transmission of a 90-byte JSON frame at 115,200 baud, 2.9 ms audio buffer), while gravity-based triggers (face detection) add a further ~91 ms from the accelerometer smoothing filter (single-pole IIR with $\alpha = 0.1$, time constant $\tau = -T/\ln(1 - \alpha) \approx 91$ ms at the 104 Hz sample rate), bringing that path to $\sim$107 ms. While this exceeds the 10 ms threshold suggested for tight rhythmic control [14], the ambient design context tolerates higher latency because gestural targets are continuous rather than discrete.

No buttons, screens, or explicit controls exist. The cube itself is the entire interface. We also note that the 6-DOF IMU is subject to gyroscope drift over sustained use; this is discussed in subsection 4.1.

2.4 Sound Design

The synthesis engine generates ambient, drone-like textures without samples. Two detuned sine-wave oscillators produce layered, slowly evolving drones that create a harmonic backdrop for the melody. The chord progression consists of three chords, which change when the cube is shaken. A triangle wave oscillator is used for melody (responding to touch). Three effect parameters respond to motion: delay time and feedback (mapped to rotation), reverb wet/dry mix (mapped to rotation and shake intensity), and low-pass filter cutoff (mapped to shake intensity). These direct mappings prioritize immediacy and discoverability over complexity, while maintaining a many-to-many approach [4] and prioritizing perceptible mapping relationships [1]. Touch events trigger melodic patterns that emerge from and recede into the ambient texture.

The choice of ambient, drone-based synthesis style reflects the intended interaction style. Slow-attack envelopes, reverb, and delay create sounds that evolve gradually, rewarding sustained gestures and continuous exploration over rapid discrete actions. This aesthetic aligns with the cube's motion-based control: tilting and rotating are inherently continuous gestures, and the synthesis engine responds in kind with smoothly morphing textures. A percussive or staccato-oriented sound design would demand precise timing and discrete gesture recognition, increasing both technical complexity and the risk of frustrating failed triggers. The ambient approach instead embraces gestural ambiguity, allowing approximate movements to produce satisfying results.

The Teensy Audio Library's block-based processing provided sufficient capability for this design, though its constraints—particularly around custom DSP—limited some sonic possibilities.

3 Preliminary Observations

SonoCube was demonstrated during class presentations at Berklee College of Music as part of an NIME-focused curriculum [12], and in informal sessions with other users, with approximately ten people interacting with the device in sessions lasting five to seven minutes each. Several observations emerged from these demonstrations. We adopt an exploratory, observation-based approach rather than controlled evaluation, consistent with frameworks that recognize designer-led reflection as a legitimate mode of NIME inquiry [7, 9].

Users consistently struggled to locate the capacitive touch surface. Of the 10 users, only 4 discovered the touch sensor within the first minute without external prompts. Two users commented that the hidden touch plate was aesthetically pleasing but difficult to find. Without a visual or tactile indication of which face responds to touch, initial exploration often missed this interaction entirely. Future iterations should incorporate subtle visual marking, such as etching or differential lighting, to indicate the touch-sensitive region.

When users did find the touch sensor, they tended to play staccato patterns, though the synthesis, with its reverb, delay, and slow attack, responds better to sustained notes. This mismatch presents a design consideration for future revisions: LED feedback could mirror the audio envelope characteristics, using slow brightness attack and decay to visually suggest sustained interaction.

Once users began moving the cube, engagement increased noticeably. The continuous sonic feedback from tilting and shaking created an exploratory loop where users tested different movements to discover their effects. One user described the interaction as "relaxing," while another noted that changing notes by rotating the cube was "fun"—suggesting the orientation-to-pitch mapping felt intuitive and playful. Users instinctively rotated the device immediately upon picking it up, but appeared hesitant to shake it vigorously enough to trigger chord changes and effect modulation.

The absence of cables or external equipment lowered barriers to initial interaction. Users could pick up the cube and immediately experience its sonic behavior without setup or explanation.

4 Discussion and Future Work

SonoCube represents a constrained design exercise. Several limitations point toward future development. A more formal evaluation, along with broader and better-structured user studies, would help identify directions for future improvements.

4.1 Limitations

The Teensy Audio Library was a limitation during the sound design and interaction design process. In the future, a board like Electrosmith's Daisy Seed could be used to increase DSP possibilities, allowing the use of pure C++ for audio implementation or MaxMSP Gen.

The IMU on board the SonoCube is currently an LSM6DSO32, which has 6 degrees of freedom. A 9-degree-of-freedom unit would add a magnetometer, providing an absolute heading reference and allowing for greater accuracy and correction of gyroscope and accelerometer drift.

4.2 Future Features

The BLE communications pipeline already exists to transmit data from the device to a computer. This could be leveraged to create a visual companion app that generates graphics based on motion, enhancing the overall experience.

Wireless communication also presents opportunities to extend the device's functionality. The existing pipeline could enable MIDI controller capabilities, allowing the cube to control external synthesizers for broader sonic possibilities and higher-quality audio output suitable for recording.

Addressing the touch discoverability issue requires design iteration through visual marking, LED indication near the touch surface, or tactile differentiation of that face. Observations and surveys will aid in finding the optimal solution.

5 Conclusion

SonoCube explores how minimal physical interfaces might support exploratory sonic interaction. Informal demonstrations suggest the approach has merit, though formal evaluation remains future work. A 70mm cube containing only motion sensing, touch detection, and ambient synthesis is intended to create an engaging exploratory experience. The dual-microcontroller architecture cleanly separates sensing from synthesis, and the self-contained form factor removes barriers to immediate engagement. While preliminary observations reveal design challenges around touch discoverability, they also suggest the appeal of continuous motion-to-sound mapping. Future work will address current limitations while exploring the cube as a platform for more sophisticated synthesis and multimodal output.

A 95-second video demo is available as Supplemental Material. It shows basic gestures and examples of the instrument's use, presenting all the basic modalities of interaction.

6 Ethical Standards

This project was developed as coursework at Berklee College of Music. No formal user studies requiring ethics approval were conducted. The observations reported are from informal demonstrations during class presentations, in which participants interacted voluntarily. No conflicts of interest exist. The project received no external funding.

Acknowledgments

The authors thank the Berklee College of Music Electronic Production and Design department for supporting this project, which was developed within the department's NIME-focused curriculum, and the members of MTEC-498 Advanced Creative Coding for their feedback during development.

References

[1] Sidney Fels, Ashley Gadd, and Axel Mulder. 2002. Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound* 7, 2 (2002), 109–126. https://doi.org/10.1017/S1355771802002042

[2] Kenya Hara. 2007. *Designing Design.* Lars Müller Publishers, Baden, Switzerland.

[3] Charles Holbrow, Elena Jessop, and Rebecca Kleinberger. 2014. Vocal Vibrations: A Multisensory Experience of the Voice. In *Proceedings of the International Conference on New Interfaces for Musical Expression.* Goldsmiths, University of London, London, United Kingdom, 431–434. https://doi.org/10.5281/zenodo.1178800

[4] Andy Hunt, Marcelo M. Wanderley, and Matthew Paradis. 2003. The Importance of Parameter Mapping in Electronic Instrument Design. *Journal of New Music Research* 32, 4 (2003), 429–440. https://doi.org/10.1076/jnmr.32.4.429.18853

[5] Fangzheng Liu, Lancelot Blanchard, Don D Haddad, and Joseph Paradiso. 2025. Two Sonification Methods for the MindCube. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME '25).* NIME, Canberra, Australia, 511–515. https://doi.org/10.5281/zenodo.15698944

[6] Romain Michon, Julius Orion Smith, Matthew Wright, Chris Chafe, John Granzow, and Ge Wang. 2017. Mobile Music, Sensors, Physical Modeling, and Digital Fabrication: Articulating the Augmented Mobile Instrument. *Applied Sciences* 7, 12 (2017), 1311. https://doi.org/10.3390/app7121311

[7] Sile O'Modhrain. 2011. A Framework for the Evaluation of Digital Musical Instruments. *Computer Music Journal* 35, 1 (03 2011), 28–42. https://doi.org/10.1162/COMJ_a_00038

[8] Bert Schiettecatte and Jean Vanderdonckt. 2008. AudioCubes: A Distributed Cube Tangible Interface Based on Interaction Range for Sound Design. In *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction (TEI '08).* Association for Computing Machinery, New York, NY, USA, 3–10. https://doi.org/10.1145/1347390.1347394

[9] Dan Stowell, Mark D. Plumbley, and Nick Bryan-Kinns. 2008. Discourse Analysis Evaluation Method for Expressive Musical Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression.* Genoa, Italy, 81–86. https://doi.org/10.5281/zenodo.1179631

[10] John Sullivan, Julian Vanasse, Catherine Guastavino, and Marcelo M. Wanderley. 2020. Reinventing the Noisebox: Designing Embedded Instruments for Active Musicians. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME '20).* NIME, Birmingham, UK, 5–10. https://doi.org/10.5281/zenodo.4813166

[11] Akito van Troyer. 2012. DrumTop: Playing with Everyday Objects. In *Proceedings of the International Conference on New Interfaces for Musical Expression.* University of Michigan, Ann Arbor, Michigan. https://doi.org/10.5281/zenodo.1178441

[12] Akito van Troyer. 2026. From Controller User to Instrument Designer: Teaching NIME in a Contemporary Music Context. In *Proceedings of the International Conference on New Interfaces for Musical Expression.* London, UK.

[13] Gili Weinberg, Maggie Orth, and Peter Russo. 2000. The embroidered musical ball: a squeezable instrument for expressive performance. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI EA '00).* Association for Computing Machinery, New York, NY, USA, 283–284. https://doi.org/10.1145/633292.633457

[14] David Wessel and Matthew Wright. 2002. Problems and Prospects for Intimate Musical Control of Computers. *Computer Music Journal* 26, 3 (09 2002), 11–22. https://doi.org/10.1162/014892602320582945