CD-Synth: a Rotating, Untethered, Digital Synthesizer

Patrick Chwalek
Responsive Environments Group
MIT Media Lab
75 Amherst Street
Cambridge, MA 02139, USA
chwalek@media.mit.edu

Joseph A. Paradiso
Responsive Environments Group
MIT Media Lab
75 Amherst Street
Cambridge, MA 02139, USA
joep@media.mit.edu

ABSTRACT

We describe the design of an untethered digital synthesizer that can be held and manipulated while broadcasting audio data to a receiving off-the-shelf Bluetooth receiver. The synthesizer allows the user to freely rotate and reorient the instrument while exploiting non-contact light sensing for a truly expressive performance. The system consists of a suite of sensors that convert rotation, orientation, touch, and user proximity into various audio filters and effects operated on preset wave tables, while offering a persistence of vision display for input visualization. This paper discusses the design of the system, including the circuit, mechanics, and software layout, as well as how this device may be incorporated into a performance.

Author Keywords

Synthesizer, multi-sensor system, lidar, rotation

CCS Concepts

•**Hardware** → **Sensor applications and deployments;** *Digital signal processing;* •**Applied computing** → **Sound and music computing;**

1. INTRODUCTION AND RELATED WORK

Rotational musical instrument interfaces are a very intuitive representation of the fundamentals of music itself, as they are inherently simple oscillators that can run from low control frequencies up to audio. Parameters like pitch (e.g. glass haramonica [13] and hurdy-gurdy [9]), tempo (e.g. Drum Buddy [12]), sample playback (e.g. CDJ [2], turntable), and delay time (e.g. tape flanging) can be easily represented and manipulated in the rotational domain. The downside of many rotating instruments is that they are bounded by the rotating mechanisms themselves, that being a fixed axle where the rotation is driven by a hand or a motor. This leads to a reduction of the performer's expressive ability to influence the generated sound by constraining them to fixed locations and/or positions. In our work, we avoid these drawbacks, but maintain the affordance of the rotational interface by incorporating non-contact sensing and wireless data transfer.

Audio synthesizers and controllers exist in many forms [11] with devices ranging in affordances to create different sounds out of a combination of oscillators, filters, and effects. For light-based synthesizer systems, transmitted light can be modulated in specific patterns to change virtually any characteristic of a sound that is produced by photodiode exposure. An example is Jacques Dudon's photosonic instrument [7, 4] from the 1980s, which was composed of a photodiode, light source, semi-transparent rotating disk, and an optical filter. The disk was patterned in such a way that when rotated, the intensity of light passing through it changes at audio frequencies (like an optical sound track on a film, or vintage optical samplers like the Optigan), and was further operated on by the handheld filter or manually moving the lightsource and/or photodetector, so that the analog signal generated by the photodiode's exposure represented the user's physical manipulations as well as the audio coming from the spinning disc. Similar to that, Mr. Quintron's Drum Buddy [12] modulated a light source by a series of slits in an encapsulating rotating canister that was sensed by a series of photodiodes positioned around the perimeter (a legacy dating to Leon Theremin's Rhythmicon). As with Dudon's instrument, the rate of rotation of the canister and light photodiode exposure could be manipulated by the performer to generate different sound effects and rhythms. However, like many rotation-based instruments, both Dudon's and Quintron's devices were stationary and physically constrained the performer. That is why we created the CD Synthesizer (CD-Synth), a mobile synthesizer which allows the user to freely rotate the instrument and manipulate the orientation while exploiting non-contact sensing for a truly expressive performance. Our device has another legacy in multimodal handheld or wearable manual controllers like Michel Waisvisz's well-known 'Hands' [14], which measure free gesture along with various kinds of finger control. Our addition of the spinning disk, however, marries these controllers with a handheld spinner, which adds an agile, tactile, mechanical low frequency oscillator (LFO) that also works as an engaging display.

Non-contact audio controllers exist in many forms and are a century old. One of the earliest and most popular is the Theremin [8], an instrument where the produced sound can be altered by the relative proximity of a user to two antennas, one controlling pitch and the other amplitude. The Theremin works by the user's proximity to the antenna changing the capacitance of the onboard oscillators which results in change of pitch and amplitude. Recent iterations of Theremin-like instruments include the Roland D-Beam controller [1], or the Termenova [10], both light-intensity-based methods that infer a user's proximity with infrared light, ranging to a maximum distance of circa 15". For our design, we modernize elements of these controllers by using a laser light detection and ranging (lidar) sensor to sense the relative distance in front of our synthesizer and translate that to pitch. This method allows us to detect

NIME'19, June 3-6, 2019, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

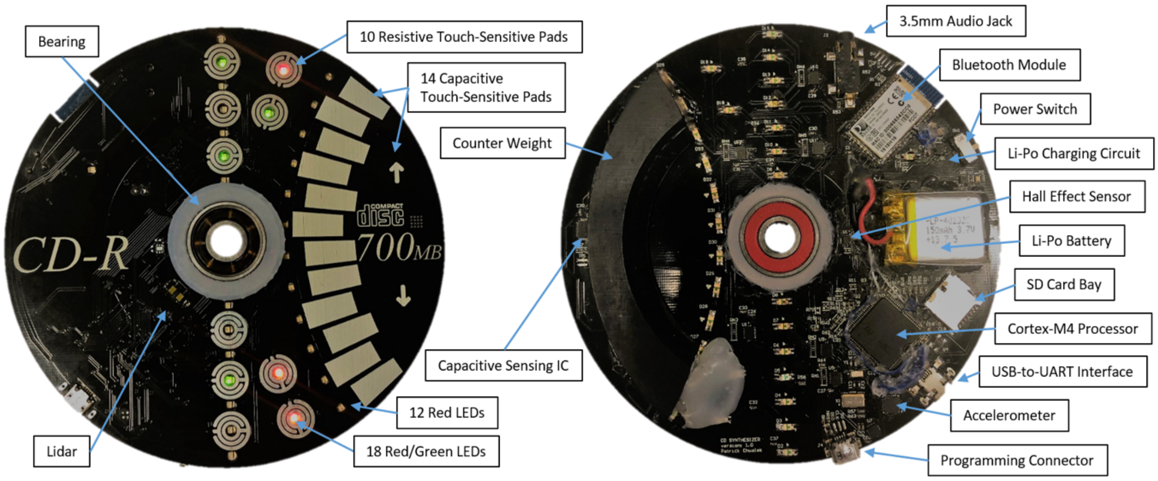

Figure 1: Front and Back Views

accurate distances of up to 4 meters (although we reduce that in software to increase signal reliability), and reduces the form factor of the overall device when compared to the Roland D-Beam. This allows the performer to be more mobile with the instrument, and to be able to freely rotate it without weight and size constraints.

Our device hosts a digital wavetable synthesizer that wirelessly transmits audio data to a nearby Bluetooth interface (e.g. speaker). The device is able to be freely manipulated by the performer and can be rotated like a fidget spinner while changing its orientation within the physical space. The device is equipped with several sensors that allows it to translate rotation, orientation, touch, and external object distance from the front face to manipulate the transmitted sounds. Additional effects such as bit crushing, low pass filtering, and wave shaping can be activated by the touch pads located on the front of the device to further customize the sound. The device is also equipped with a persistence of vision display that when rotated, shows visualizations that can be mapped to the user's input.

2. DESIGN

The system was designed to appear like a compact disc, with the only similarities being size, shape, and textual features (suggesting packaging with a CD as custom merchandise, for example). The three main components of the system are the single circuit board, mechanical components that include the handle and bearing, and the embedded software which includes sensor handling and digital signal processing.

2.1 Hardware

The circuit consists of an ARM Cortex-M4 processor that is capable of running at 120MHz, with 2MB of included program memory. This specific processor was selected since the intended application is for wavetable synthesis and the large tables can be saved into program memory instead of relying on external storage, simplifying the overall design. The selected processor also includes a floating-point unit (FPU) and digital signal processing (DSP) instructions, which significantly decrease the processor cycles required for any operations on the output data stream, resulting in higher possible sampling rates of the system's sensors and a reduction in user input latency.

The system has a triple-axis accelerometer (MMA8451), a lidar sensor (VL53L1X), Hall effect sensor (AH8503), and a capacitive sensing integrated circuit (CAP1214), which are all used for user input. Separating the sensors from the voltage regulator is a complementary metal-oxide-semiconductor (CMOS) switch that can be turned on or off by the processor to conserve battery power when not active. For capacitive sensing, there are twelve keys, as shown in Figure 1, that are preprogrammed to playback a selected wavetable at any of the twelve notes within a selected octave. An octave can be incremented or decremented by touching the up or down arrow capacitive keys, respectively. There are also 10 resistive touch-sensitive areas on the disk that are used for configuring the system (i.e. wavetable selection, filter characteristic changes, sensor activations, etc.). The keys are left purposely unlabeled so that they may be reconfigured to a user's preference if developing on the device.

For visualizations, there are twelve red LEDs that correspond to each of the twelve capacitive keys to give feedback to the user that a pad has been touched. In addition, there are ten red/green LEDs for the resistive touch-sensitive keys so as to increase the dimension of color representation to a user's input for multiple feature selection (e.g. green, orange, red). There are eight additional red/green LEDs that are used primarily for visualizing the user's distance from the lidar sensor and to increase the resolution of the persistence of vision display that is activated when the disk is rotated. To quickly toggle each of these LEDs, six eight-channel shift registers were connected serially to allow for all 48 LEDs to be manipulated while only using three pins on the microcontroller. The brightness of all the LEDs can be varied by changing the duty cycle of the PWM signal that is transmitted to the shift registers.

The system is designed to be easily programmable so it has the serial wire debugging (SWD) interface of the controller broken out into a Micro USB connector. There is also a USB-to-UART bridge that can be connected via another, separate Micro USB connector, that allows for debugging and interfacing with other machines when the device is not rotating. The system also includes a lithium polymer charging circuit so when interfacing with one or both of these connectors, the onboard battery will also be charging through the 5V USB cable. Furthermore, there is a Micro SD card bay on the device that allows for the option of quickly swapping in custom wavetables.

The radio communication link uses an RN-52 Bluetooth radio module. This chip was picked specifically for its support of the Bluetooth Advanced Audio Distribution Profile (A2DP) and Audio/Video Remote Control Profile (AVRCP) that allows it to pair with any Bluetooth speaker. The intention was to use the Integrated Inter-IC Sound Bus (I2S) inputs of the device to digitally transmit the audio from the controller to the module, but the factory-installed firmware on the module has the feature deactivated. As a solution, the analog inputs were used with the digital-to-analog converter (DAC) of the controller to transmit the audio. However, this results in signal resolution loss due to the unnecessary data conversions so future work is warranted in finding a more suitable module. In principle, the sensor data could be transmitted via the Bluetooth link as well, allowing this device to be used as a controller for more complex external synthesis or processing.

To hold the device, a handle assembly (Figure 2) is designed, which includes a shielding disk to keep the user from hitting any of the electronics while the disk is spinning. The device is designed to rotate about its center, as a compact disk would in a disk drive. Several bearings were tested, but the most optimal one was found to be a standard 608 size bearing that is commonly used in fidget spinners. The shielding disk has 12 embedded magnets (Figure 3) for the Hall effect sensor to sense the rotational rate of the system.

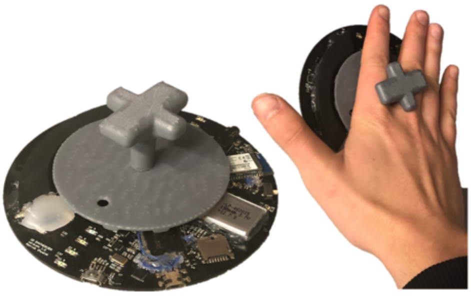

Figure 2: Handle and Disk Attached and Held

Figure 3: Magnets in Shielding Disk (only 4 shown)

2.2 Software and Control

The controller outputs 12-bit audio at 40kHz to the Bluetooth module while also communicating with all the sensors, processing the data, and preparing the next packet of audio data to be transmitted. To best manage all the individual tasks, FreeRTOS, an open source real time operating system, is used. FreeRTOS allows for the abstraction of tasks and priority managing so that a time-sensitive operation, such as preparing the next packet of data, can preempt any currently non-critical running operations. As is shown in Figure 4, there are five threads that deal with the sensors, each running at their own sample rates. The highest priority thread is the audio synthesis thread, and that is unblocked each time the DAC requests the next buffer of data. To increase the efficiency of the system, the DAC peripheral accesses each buffer of data using direct memory access (DMA) so that the processor does not have to dedicate cycles on handling that operation.

In the system's current configuration, the resistive touch-sensitive pads allow for the control of wavetable selection (sine, saw, and triangle waves) and wave shape activation and selection (sigmoid, hyperbolic tangent, and 4th-order Chebyshev polynomial [5]). The lidar and capacitive keys can also be toggled, bit crushing can be enabled and varied, low-pass filter activated, and the filter's resonance (Q factor) can be modified to alter the bandwidth and damping of the filter. When the distance sensing is enabled, the frequency of the output is varied depending on how far the user's hand, or other object, is positioned from the top of the device. The programmed range is 1.3 meters and the frequency-to-range mapping is logarithmic to better match with human perception of relative loudness [15], with higher frequencies mapped to longer distances. When the distance exceeds 1.3 meters, the sound turns off.

A second-order infinite impulse response (IIR) biquad form 2 digital filter [6] is implemented and can be enabled with a touch pad press. The cutoff frequency of the filter is mapped to the orientation of the device, increasing logarithmically as the device is turned upright. The orientation of the device is calculated after a low pass filter is applied on the accelerometer data, removing any artifacts from rotation or movement, and comparing the resultant vector with the gravity vector. Since the cutoff frequency is changed in real-time, the coefficients of the filter are calculated at each accelerometer sampling (20Hz). As mentioned before, the Q factor can be incremented in real-time to four preset values, changing the frequency response of the filter.

The system has a row of 14 red/green LEDs that are positioned along a radial axis. When rotating, magnets pass the Hall effect sensor which trigger a hardware interrupt based on a programmable comparator threshold. If the time interval between each hardware interrupt is small enough (i.e. disk is rapidly spinning), the POV display is triggered and the LEDs are modulated to the rate of the spin to form a 2D image. The current implementation (Figure 5) cycles through red and green to represent lidar sensed distance, spinning clockwise when moving further from the device and counterclockwise when moving toward.

There is also a 7-second record feature on the device that allows for recording and repetitive playback on the device. This time duration is limited to the onboard available memory but future iterations can have the recording save to the SD card to greatly increase the duration. Future work will be in expanding the SD card functionality so that custom wavetables can be easily imported while also giving the ability to record the entire performance.

3. PERFORMANCE

In [3], several features of the device are shown. When the disk is held with one hand (Figure 2), the device can be initially configured using the resistive touch pads to choose the desired wave table, enabling the filtering, modifying the filter's resonance (Q factor), choosing the waveshape, and enabling either the lidar or capacitive keys. If the capacitive keys are selected, the twelve keys can be played and mixed together as the performer chooses. To record sounds for immediate playback, the recording can be activated, deactivated, and played back by a series of presses on the designated recording touch pad.

When the lidar is enabled, the user can place their hand over the device and a sound will be generated based on their hand's distance relative to the front-facing lidar. If quickly moving their hand in and out of the lidar's field of view (FOV), a 'record scratching' sound effect is heard. Given that the lidar is not constrained to just measuring hand distance, the disk can be played off of any visibly opaque object. When a more transparent object, specifically in the

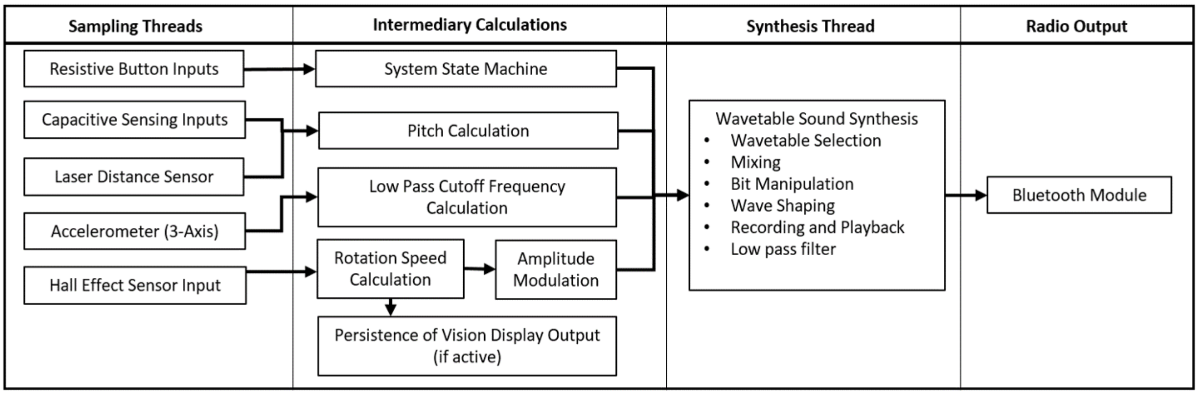

Figure 4: Software Flow Diagram

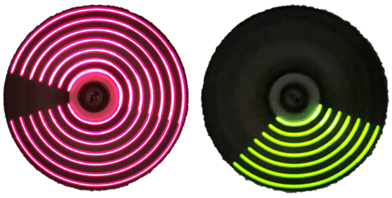

Figure 5: Two POV states taken at long camera exposure

940nm domain, is used as the distance source, the distance measurements increases in error which results in frequency jitter.

When the disk is spun, the amplitude of the generated sound is modulated at the rate of the spin, increasing as the device is rotated faster. The POV display is automatically toggled when sudden rotation is detected and the resultant visualization is mapped to the sensor readings of the lidar. From a stationary position, the disk can be turned slightly to slowly increase or decrease the sound.

4. CONCLUSION AND FUTURE WORK

In this paper, we have introduced the CD-Synth, a rotating digital synthesizer that is untethered and can be freely manipulated to sculpt the transmitted sounds. The synthesizer allows the user to freely rotate and reorient the instrument while exploiting non-contact light sensing for a truly expressive performance. The device takes the form factor of a compact disk and is equipped with a suite of sensors to sense any type of movement of the device. In addition, the system includes a persistence of vision display that is activated while the device is rotated to visualize the user's variable input. With the addition of an SD card, a performer can add their own wave tables and settings to further customize their sound.

Future improvements include replacing the selected Bluetooth module with an aptX version that further reduces any latency between the performer's actions and the sound output. It is also desired to move to a Bluetooth module that has I2S support to mitigate any unnecessary conversions. In addition, a radial array of lidar sensors would increase the sensor feature space for more granularity of the distance sensing from the front of the device which could sense lateral topological changes. This could lead to interesting effects when the user tilts the object that is being measured to form different sound effects. Finally, adding a digital microphone would allow for recording of external sounds that could be operated on and played back for added flexibility.

5. ACKNOWLEDGMENTS

We would like to thank Mark Feldmeier and his Audio Electronics course at the MIT Media Lab that aided in the initial conception of the CD-Synth. We would also like to thank Spencer Russell and Brian Mayton in the Responsive Environments group for their discussions and guidance in the digital signal processing effort.

6. REFERENCES

[1] Roland d-beam. `https://www.roland.com/global/products/mc-505/`, December 2007. Last accessed 20 January 2019.

[2] Industry standards: Pioneer CDJ. `https://www.residentadvisor.net/features/1910`, October 2013. Last accessed 24 January 2019.

[3] CD-Synth Demo. `https://vimeo.com/313665367`, January 2019. Last accessed 27 January 2019.

[4] D. Arfib and J. Dudon. A digital version of the photosonic instrument. In *ICMC*, 1999.

[5] T. Barrass. Mozzi: Sound synthesis library for arduino. `https://github.com/sensorium/Mozzi`, January 2019. Last accessed 20 January 2019.

[6] R. Bristow-Johnson. Cookbook formulae for audio eq biquad filter coefficients. `http://www.musicdsp.org/files/Audio-EQ-Cookbook.txt`. Last accessed 20 January 2019.

[7] J. Dudon and D. Arfib. Photosonic disk: interactions between graphic research and gestural controls. 2000.

[8] A. Glinsky. *The Theremin in the emergence of electronic music.* ProQuest NYU, 1992.

[9] R. A. GREEN. *The Hurdy-Gurdy in Eighteenth-Century France, Second Edition.* Indiana University Press, 2 edition, 2016.

[10] L. Hasan, N. Yu, and J. Paradiso. The termenova: A hybrid free-gesture interface. pages 122–127, 05 2002.

[11] J. A. Paradiso. Electronic music: new ways to play. *IEEE Spectrum*, 34(12):18–30, Dec 1997.

[12] Quintron. The drum buddy. `https://www.youtube.com/watch?v=3rt-MUx_NLY`, December 2007. Last accessed 20 January 2019.

[13] T. D. Rossing. Acoustics of the glass harmonica. *The Journal of the Acoustical Society of America*, 95(2):1106–1111, 1994.

[14] G. Torre, K. Andersen, and F. BaldÃľ. The hands: The making of a digital musical instrument. *Computer Music Journal*, 40:22–34, 06 2016.

[15] L. R. Varshney and J. Z. Sun. Why do we perceive logarithmically? *Significance*, 10(1):28–31.