Noisebox: Design and Prototype of a New Digital Musical Instrument

John Sullivan
Input Devices and Music Interaction Laboratory, McGill University
`john.sullivan2@mail.mcgill.ca`

ABSTRACT

This paper documents the initial prototyping of a new digital musical instrument. Specifically, it focuses on design of the interface, and contextualizes the project through some of the existing research in the field of gestural control of new musical instruments. The project began with a concept for a standalone hand-held polyphonic synthesizer called the Noisebox (Figure 1). Several key concepts and strategies were explored and implemented during its development, including: analysis and application of gesture in musical performance, choice of sensors and sensor conditioning, appropriate mapping strategies, and evaluation of user experience. The outcome yielded a functional prototype that fulfilled the initial goal of the project to design and build a working instrument from start to finish. The stage documented here represents the first phase of a longer project. Future phases will conduct user tests to measure the success of the instrument based on performer feedback and refine the design through multiple iterations, leading to a finished instrument.

1. INTRODUCTION

This project began as a way to apply fundamental concepts of designing input devices for new musical instruments directly to practice. A new instrument called the *Noisebox* [1] was conceived and built to test the capabilities of the Raspberry Pi as a platform for low cost, embeddable processors for digital musical instruments (DMIs). The design attempted to embody some of the characteristics of analog instruments, most importantly of reuniting the controls and sound production together in one discrete unit. This feature marks a reversal of a primary characteristic of DMIs, where the lack of acoustical coupling of physical control and sound production has allowed for complete separation of these two systems [1] [2]. Other strategies included removal of external wires and connections to auxiliary components, and a focus on simple, learnable controls. The Noisebox is intended to be easily held and manipulated in the hands of a performer.

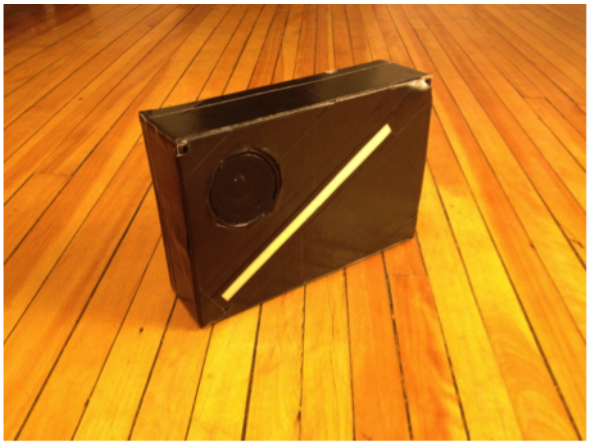

Figure 1. First prototype of the Noisebox.

2. DESIGNING THE NOISEBOX

The process of designing and building the instrument spanned four sections. First, a summary study of gesture was used to plan a basic control system that would be inherently intuitive and playable for a performer. Then sensors and sensing strategies were chosen and implemented. A polyphonic FM synthesizer was programmed in the visual programming language Pure Data [2] . Finally, a software mapping system was devised to connect the performer's gestures to sound production.

2.1 Gesture Selection

When considering how a performer might interact with the proposed instrument, priority was given to create a set of controls that would be simple and intuitive. Direct gesture acquisition was chosen over indirect or physiological methods as this offered the most straightforward connection between performer and instrument [3]. Studies have shown that a tight coupling between performer and instrument is a key factor in achieving musical expression [1]. This relationship is linked to the perception of expressiveness by both performer and audience [4] [5]. Inspiration was drawn from the relationship between a skilled performer and acoustic instrument, in which the instrument has been described as an extension of the musician's body [6].

---

[1] Video demonstration: http://vimeo.com/113886990

[2] This paper focuses specifically on interface design, and sound synthesis is not covered here in depth.

Cadoz and Ward [7], Delalande [8] and Dahl et al. [9] offer similar classifications of levels of gesture, from functional (sound-producing) to symbolic (non-sound-producing). Using Delalande's classification, the primary mode of gesture for the Noisebox is *effective*, using handed gestures of tapping and sliding across a specially designated surface. Another class of control is available, that can be classified as an *accompanying* gesture. This is achieved through manipulation and orientation of the instrument through physical space. The effective gestures of tapping and sliding to control sound parameters closely mimics controls of many traditional analog instruments. Movement of the instrument in physical space is also common with traditional instruments, however the production or modulation of sound is uncommon. With the Noisebox, these gestures add a wide array of sound parameters that the performer can control. This demonstrates the use of effective, or *ancillary* gesture that can be used to extend musical control beyond the normal capabilities of a traditional acoustic instrument [10].

2.2 Sensors and Signal Acquisition

With the methods and types of gestures established, the next step was to select the appropriate sensors and technology to acquire the gestural data. Two types of data needed to be captured – continuous variables, and discrete, event-based signals, which Max Mathews referred to as *triggers* [11].

Piezoelectric sensors were selected to capture the discrete signals. To improve their accuracy, signal conditioning was applied through software to set appropriate thresholds and prevent unintentional triggering. These sensors were chosen for their low cost and simplicity, however during testing we found that other sensors would have been a better choice. This use of "unsophisticated engineering solutions" [12] has been identified as a common but troublesome trend in DMI design. Though far more robust technologies exist, they are often prohibitively expensive and require an advanced level of expertise to implement. However, use of cheaper and lower tech solutions (as with our piezo sensors) comes at the cost of reduced accuracy and precision in the gesture acquisition.

A SoftPot linear position sensor was used to capture the sliding gesture. This can function as either an event-based or continuous control, where a single value can be specified by a single touch, or continuous values can be sent with a continuous motion. Again, conditioning was applied through software to attenuate the input signal to a suitable range and to freeze values at their last position until further modulated. Other sensors were considered and may be substituted in future iterations. One promising alternative is the use of force sensors made of conductive paper [13].

A study by Marshall et al. [14] on performer preference of input gesture found preference for pitch selection by a "pressing" gesture (i.e. use of buttons or keys) over "sliding" gestures. While this suggests that our instrument might benefit from a different mode of input for pitch selection, we found that the sliding control worked quite well, especially for glissando type pitch modulations, for which the Noisebox is well-suited.

Finally, to capture the physical manipulation of the instrument, the MPU-6050 accelerometer-gyroscope sensor was used. Accelerometers and inertial measurement unit (IMU) sensors are among the most widely used sensors in DMIs today [12]. Some IMU sensors also integrate a magnetometer, which orient an object in the physical world by measuring the Earth's magnetic field [15]. The MPU-6050 lacks a magnetometer, so instead algorithms were programmed to "zero out" the instrument's physical orientation over time to keep the performer's controls consistent and predictable.

Sensor fusion for the MPU-6050 is contained onboard the sensor's integrated circuit firmware [3] . Accelerometer and gyroscope data correlated to provide highly accurate measurements of three axes: yaw, pitch and roll. Additional signal conditioning was applied to the continuous data stream to limit the sampling rate of the sensor to 50Hz. This was found to be high enough to be extremely responsive while sufficiently limiting the bandwidth to an acceptable range for the serial communication protocol that connected the sensors to the Raspberry Pi.

2.3 Mapping

Mapping objectives were laid out to create an instrument that could adhere to Wessel and Wright's principle of "low entry fee with no ceiling on virtuosity" [16]. This meant setting up simple and intuitive controls that could easily be understood and interpreted by a novice while containing sufficient nuance and complexity to reward continued practice with greater expression. The objectives were achieved by implementing one-to-one mappings for some parameters like turning individual voices on and off, and many-to-one and many-to-many mappings for frequency and timbral control of the sound synthesis [12]. Inspiration was taken from Wessel's research on timbre space for musical control [17] for higher-level parameters of overall sound output. In practice this was approached by creating two levels of control – first on a low-level, voice-by-voice basis where frequency of each voice can be controlled discretely and number of simultaneous voices can be controlled, and second (high-level) by modulation and depth parameters of the FM synthesis that can be applied to all active voices simultaneously [18].

Consistent with research by Hunt and Kirk [19], Hunt, et al. [20], and Kvifte [21], more complex mappings were ultimately the most rewarding and engaging from a performance perspective. Based on preliminary user testing, the instrument was most effective when the individual low-level controls were shifted out of focus and the performer began to work intuitively, shaping the timbral characteristics of the overall sound output. This intuitive mode of performance also reinforces the benefits of tight coupling between performer and instrument and in turn, the coupling of interface and sound production.

One of the biggest technical challenges to achieving this tight coupling is achieving sufficiently low latency between

---

[3] MotionFusion^{TM}, by InvenSense: http://www.invensense.com/

gesture and sound. Wessel and Wright suggest acceptable latency thresholds of less than 10ms with a range of variation no more than 1ms [16]. So far the Noisebox hasn't come close to this for a variety of reasons. One is the limitations of the Raspberry Pi Model B, with 512MB of RAM and 700MHz processor speed. Additionally, the synthesis and mapping algorithms could be rewritten to optimize performance. However, the instrument – which produces sustained legato tones and has been augmented with reverb and delay to create a lush, ambient sound – is somewhat forgiving in this regard.

One exceptional mapping strategy employed in the Noisebox is the voice selection algorithm. The performer is able to activate up to eight simultaneous voices. Once multiple voices are in play, the performer is able to 'select' control of any single voice by orienting the instrument across a 180 degree horizontal plane. Thus, aiming the Noisebox to the performer's far left activates primary control of the first voice, and moving the device across the body to the performer's right side sequentially selects control of each individual voice up to the last. The pitch, timbre and loudness of each voice can be modulated. While the voice selector is a discrete control, as previously mentioned, the instrument is most effective when the performer shifts focus from low-level concern of individual voices to higher-level control of timbral space.

3. USER EXPERIENCE EVALUATION

The evaluation of user experience throughout the process is important to inform the design and assess its success. More comprehensive testing and analysis is planned in future phases that will fully guide the development of the instrument.

Several components of the design were implemented with the end user in mind. The main objective of the building the Noisebox was to create an instrument that would be accessible, interesting and enjoyable for a performer. The aesthetic design was intended remove the DMI and its user from typical performance configurations – for example, the performer hunched over a laptop or tethered to wires and auxiliary equipment. This was implemented by building a completely stand-alone instrument.

As an interface for control of sound, we tried to strike a balance between what Michel Waisvisz referred to as a "meager recreation of existing concepts and imitation of analogue worlds" [11] and the unchecked potential of computer-based instruments, described by Atau Tanaka as a "theme park one-man-band" [22]. This was carried out by using some of the aesthetic qualities and characteristics of acoustic instruments (familiar gestures, direct control over primary sound variables) while exploring enhanced capabilities available exclusively in the digital realm (acquisition of ancillary gestures, certain complex mappings).

Though not addressed in depth here, adequate feedback is an important and complex topic, and is vital to creating a successful user experience [11]. The primary channel of feedback for the Noisebox is auditory. A secondary source is vibrotactile, conveniently present thanks to the sound production embedded within the instrument itself. The housing of the instrument creates a natural resonance chamber that provides significant haptic feedback. This is another way in which the Noisebox borrows from its acoustic counterparts.

Ultimately, the true measure of successful user experience will be demonstrated by continued use and adoption into use by multiple users. This is a challenge for all designers of DMIs and may not always have to do with technical utility or usability of an instrument. Wessel and Wright suggest that instruments and interfaces succeed for mostly sociological reasons [16]. It seems that there is a general consensus though, that successful instrument and interface design achieves an optimal balance of engineering technology and musical sensibility. While still it its early development, the Noisebox shows promise in these areas. An important next step is to begin dedicated user evaluation to collect and analyze data for further development and refinement.

4. CONCLUSION

This paper has summarized the process of designing and building a novel input device for a new digital musical instrument and placed it in the context of current and previous interdisciplinary research in the technical and creative fields of human-computer interaction, computer and electrical engineering, design, art and music performance. Consideration of these areas guided design of the Noisebox through the selection of gestures for instrument control, sensors and mapping strategies. User experience design was utilized to create an instrument that was specifically tailored to be functional and engaging for the performer and to encourage lasting and repeated use.

The current version of the Noisebox is an initial prototype. Future stages will refine the working model with the ultimate goal of producing a finished family of instruments. Throughout the process of designing and building, several areas have been identified to improve upon or redesign, including the separation of mapping layers into one or more discrete modules and refining of gesture acquisition data with better sensor technologies and circuit conditioning techniques.

Other important aspects of this project were not covered in this paper but are integral nonetheless and demonstrate areas for further research. Sound synthesis was achieved though a low bandwidth polyphonic FM synthesizer programmed in Pure Data. Improvement and optimization of synthesis algorithms and code is necessary to lower latency and improve overall performance and sound quality. The permanent physical construction of the body of the instrument has been designed but not constructed, and will contribute significantly to the instrument as a whole. Use of the Raspberry Pi Model B has revealed limitations for processing the bandwidth necessary to sample sensor data at sufficiently high rates and to perform advanced digital signal processing. Experimentation with the newer Raspberry Pi 2 and other development boards like the BeagleBone Black and Intel Galileo will likely provide better results. Finally, implementation of feedback requires dedicated attention to ensure that sufficient responsiveness is available for the performer.

Acknowledgments

The author would like to thank the Input Devices and Music Interaction Laboratory and director Marcelo Wanderley at McGill University, and Harvestworks Digital Media Art Center in New York for support and guidance on this project. Special thanks to the University of Maine Intermedia MFA program and the Graduate Student Government, whose funding and financial support made this research possible.

5. REFERENCES

[1] A. Mulder, "Towards a choice of gestural constraints for instrumental performers," in *Trends in gestural control of music*, M. Wanderley and M. Battier, Eds. Ircam - Centre Pompidou, 2000, pp. 315–335.

[2] B. Bongers, "Physical Interfaces in the Electronic Arts Interaction Theory and Interfacing Techniques for Real-time Performance," in *Trends in Gestural Control of Music*, 2000, vol. 2000, pp. 41–70.

[3] M. M. M. Wanderley and P. Depalle, "Gestural control of sound synthesis," *Proceedings of the IEEE*, vol. 92, no. 4, pp. 632–644, Apr. 2004.

[4] S. Fels, A. Gadd, and A. Mulder, "Mapping transparency through metaphor: towards more expressive musical instruments," *Organised Sound*, vol. 7, no. 02, pp. 1–45, Jan. 2003.

[5] J.-W. W. Davidson, "Visual Perception of Performance Manner in the Movements of Solo Musicians," in *Psychology of Music*, vol. 21, no. 2, Apr. 1993, pp. 103–113.

[6] R. Vertegaal, T. Ungvary, and M. Kieslinger, "Towards a Musicians Cockpit: Transducers, Feedback and Musical Function," in *ICMC'96*, no. criterion 2, 1996, pp. 308–311.

[7] C. Cadoz and M. Wanderley, "Gesture-music," *Trends in Gestural Control of Music*, pp. 71–94, 2000.

[8] A. R. Jensenius, M. M. Wanderley, R. I. Godoy, and M. Leman, "Musical gestures: Concepts and methods in research," in *Musical gestures: Sound, movement, and meaning*, 2010, pp. 12–35.

[9] S. Dahl, F. Bevilacqua, R. Bresin, M. Clayton, L. Leante, I. Poggi, and N. Rasamimanana, "Musical gestures: Gestures in Performance," in *Musical gestures: Sound, movement, and meaning*, R. I. Goldø y and M. Leman, Eds. Routledge, 2010, ch. 3, pp. 36 – 68.

[10] M. M. Wanderley, B. W. Vines, N. Middleton, C. McKay, and W. Hatch, "The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field," *Journal of New Music Research*, vol. 34, no. 1, pp. 97–113, Mar. 2005.

[11] M. Battier, M. Wanderley, and J. Rovan, "Round table: Electronic Controllers in Music Performance and Composition," in *Trends in Gestural Control of Music*, 2000, pp. 415–438.

[12] C. B. Medeiros and M. M. Wanderley, "A comprehensive review of sensors and instrumentation methods in devices for musical expression." *Sensors (Basel, Switzerland)*, vol. 14, no. 8, pp. 13 556–91, Jan. 2014.

[13] R. Koehly, M. M. Wanderley, T. van de Ven, and D. Curtil, "In-House Development of Paper Force Sensors for Musical Applications," *Computer Music Journal*, vol. 38, no. 2, pp. 22–35, Jun. 2014.

[14] M. T. Marshall, M. Hartshorn, M. M. Wanderley, and D. J. Levitin, "Sensor Choice for Parameter Modulations in Digital Musical Instruments: Empirical Evidence from Pitch Modulation," *Journal of New Music Research*, vol. 38, no. 3, pp. 241–253, Dec. 2009.

[15] S. Macintyre, "Magnetic Field Measurement," in *The Measurement, Instrumentation and Sensors Handbook on CD-ROM*. CRC PRess, 1999.

[16] D. Wessel and M. Wright, "Problems and Prospects for Intimate Musical Control of Computers," *Computer Music journal*, vol. 26, no. 3, pp. 11–14, 2002.

[17] D. Wessel, "Timbre Space as a Musical Control Structure," *Computer Music Journal*, vol. 3, no. 2, pp. 45–52, 1979.

[18] V. Verfaille, M. M. Wanderley, and P. Depalle, "Mapping strategies for gestural and adaptive control of digital audio effects," *Journal of New Music Research*, vol. 35, no. 1, pp. 71–93, Mar. 2006.

[19] A. Hunt and R. Kirk, "Mapping Strategies for Musical Performance," in *Trends in Gestural Control of Music*, M. Wanderley and M. Battier, Eds. Paris: IRCAM - Centre Pompidou, 2000, pp. 231–258.

[20] A. Hunt, M. M. Wanderley, R. Kirk, I. C. Pompidou, and P. France, "Towards a Model for Instrumental Mapping in Expert Musical Interaction University of York Analysis-Synthesis Team," *Proc. of the Int. Computer Music Conf. (ICMC'2000), ICMA: Berlin, Germany.*, 2000.

[21] T. Kvifte, "On the Description of Mapping Structures," *Journal of New Music Research*, vol. 37, no. 4, pp. 353–362, Dec. 2008.

[22] A. Tanaka, "Musical Performance Practice on Sensor-based Instruments," in *Trends in Gestural Control*, M. Wanderley and M. Battier, Eds. Centre Pompidou: Paris: IRCAM, 2000, no. Laurel 1990, pp. 389–406.