SONIMIME: SONIFICATION OF FINE MOTOR SKILLS

Jesse Fox, Jennifer Carlile, Jonathan Berger

CCRMA
Stanford University
Stanford, California
USA

jrobfox@ccrma.stanford.edu
jcarlile@ccrma.stanford.edu
brg@ccrma.stanford.edu

ABSTRACT

This paper describes the design of SoniMime, a system for the sonification of hand motion. Among SoniMime's applications is the use of auditory feedback to refine fine motor skills in a wide variety of tasks. Our primary sonification method involves mapping movement to timbre parameters. Specifically, we explore the application of the tristimulus timbre model for the sonification of gestural data, working toward the goals of assisting a user to learn a particular motion or gesture with minimal deviation. We also explore real-time timbre shaping and its possible use as an intuitive sonification tool related to formant synthesis and human vowel recognition. SoniMime uses two 3-D accelerometers connected to an Atmel microprocessor which outputs OSC control messages. Data filtering, parameter mapping, and sound synthesis take place in Pd running on a Linux computer.

1. INTRODUCTION

Imagine learning to control a cello bow without the auditory feedback of the bowed string. Next, imagine an analogous auditory haptic feedback loop applied to learning a delicate tai chi pattern. In this paper we describe SoniMime, a system that provides auditory response to fine motor movement of the hand with the purpose of refining controlled hand motion.

1.1. Description

SoniMime uses a pair of 3-D accelerometers to track hand gestures, while an Atmel microprocessor formats the serial accelerometer output into Open Sound Control (OSC) messages. All data filtering and sound synthesis is implemented using Pure Data (Pd) running on a Linux computer. Due to the precision and flexibility of the sensors used, we can deduce several different types of gestural movement, and thus several different schemes for the sonification of gesture data are explored.

In general, it is found that direct mappings of tilt and jerk to various synthesis parameters are more successful than memory-based, comparative pattern-matching schemes. One implementation of SoniMime involves a gesturally-controlled tristimulus timbre synthesizer [1] [2], resulting in the generation of speech-like vowel formants.

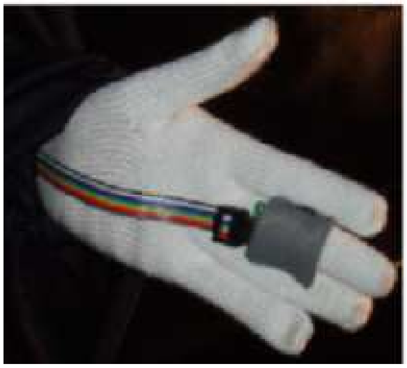

Figure 1: 3-D accelerometer used to sense movement

1.2. Motivation

SoniMime was originally conceived as a performance tool for dancers. When dancers are given the ability to create sound and contribute to the overall sonic environment, choreography and composition are suddenly infused with broadened horizons. Further, when dancers have an intuitive, reactive system at their disposal, the lines separating dancer, composer, musician, and even audience are blurred. It is for the development of this sort of reactive system that SoniMime was originally devised.

SoniMime provides a means for a person to physically explore a real space through auditory feedback, as in the case of a dancer or athlete. SoniMime senses both fine and gross motor movements, ranging from the slightest tilt or tap of the finger to a large sweep of the arm, actively engaging the user in the sonification algorithm.

In sonification systems, Fernström et al [3] emphasize the importance of continuous interaction with the sonification algorithm. Continuously mapping high resolution movement data to recognizably change sound synthesis parameters actively engages the user in the auditory feedback loop. SoniMime's resolution and responsiveness allows a user to correlate even the smallest of one's own actions with auditory feedback in real-time. We highlight the low latency of our system, as it concretizes the relation between the user's actions and the system's reactions, facilitating learning and engaging the user.

2. SYSTEM

SoniMime uses a pair of ADXL ADC digital accelerometer boards produced by Procyon Engineering [4]. Each board houses a pair of bidirectional accelerometers that, when mounted perpendicularly, sense acceleration in 3 dimensions. The change in voltage associated with acceleration in any direction is sent through an A/D converter that outputs a single serial data stream based on the I2C data protocol. This data stream is interpreted by an Atmel ATMega16 microprocessor that outputs OSC messages via serial port. A computer running Pure Data (Pd) under Linux houses all of our data filtering, mapping and sound synthesis software. According to Levitin, for an association to exist between physical motion and sonic result, the latency between manual input and auditory output must be less than 10 msecs, or perceptually zero [5], so low latency was a high priority in SoniMime's system design.

The physical design of the system is meant to be small and unobtrusive, attributes important to a dancer or other potential user. Each accelerometer board is attached to the hand via an elastic band [see fig. 1] and a small foam pouch, to ensure user comfort. Wires run up the sleeves to an AVRmini development board [4] that is held in a belt-pouch, and powered using a 9V battery. The AVRmini board houses our microprocessor, and a serial cable directly connects the board to the Linux computer.

All data is sent to Pd as a single 30-bit stream of OSC messages, running at 115200 bits per second. Within Pd, each 30-bit number is split into its 3x10-bit directional components (x, y, and z). DC offset is removed, and the data is filtered in three ways: First, a series of high-pass filters output accelerometer data corresponding to sudden jerks or impacts. Second, a series of low-pass filters output accelerometer data corresponding to tilt. Third, tilt data is differentiated (using a one-sample memory buffer) to output jerk, or change in acceleration. Our final OSC data stream then consists of three separate outputs for each direction, or nine outputs total for each hand. Finally, a "stillness detector" is implemented that changes state based upon continual analysis of tilt data from each hand. The end result is a Pd abstraction that has a total of twenty outputs (9 tilt/acceleration signals, plus one "stillness detector" for each hand).

3. SONIFICATION AND CONTROL

SoniMime translates acceleration data with high resolution and low latency into a data stream that reflects three specific kinds of hand movement–tilt, jerk, and impact. Due to the large number of movement parameters we are able to sense, SoniMime has potential applications in the exploration and sonification of high-dimensional data sets. Also, this allows us to easily adapt SoniMime for use in a wide array of musical applications, such as sound spatialization, playback control, mixing, and synthesis control.

Through extensive experimentation, it was found that the use of pattern-matching schema in pursuit of real-time gesture recognition often fell short of our original ideal of an intuitive, reactive system. Rather, we chose to focus on the sonification of the true output of our sensors. By mapping various acceleration parameters directly to sonic parameters such as pitch, amplitude, playback rate, or virtual location in a timbre space, we found the result at the user's end to be much more pleasing. Through simplified mapping of sensor data, we could focus on low latency and reactivity, two features that make interacting with SoniMime so pleasing and intuitive.

With a reactive system in place, our task shifted to best utilizing these sensor data mapping methods in devising aurally pleasing systems for the sonification of movement. Many methods were implemented, ranging from application-specific paradigms (learned motions through auditory feedback) to more general paradigms (such as the tristimulus timbre synthesizer described in section 3.1).

The complete SoniMime software package (including the tristimulus timbre synthesizer patch discussed in section 3.1 and links to musical applications) is available online [6].

3.1. Tristimulus Timbre Model

An interesting application of SoniMime is a physical implementation of the tristimulus timbre model for sound shaping. In Pd, we created a patch that maps sensor data to control frequency, amplitude, and timbre of a synthesized sound.

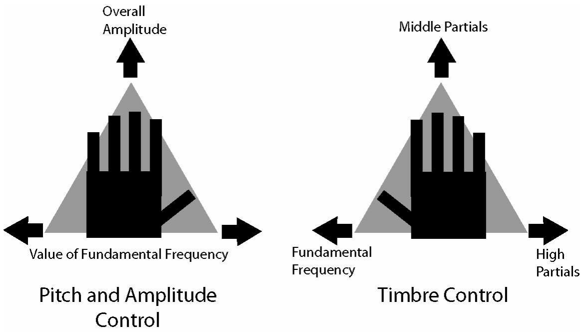

Figure 2: Parameter Mapping of X and Z tilt

X and Z tilt data from the "timbre" hand (which could be either the left or right hand) determine the weight given to each of the three timbre stimuli: the fundamental frequency, the middle partials (2-5), and the high partials (6-16). The sum of the weight given to all three stimuli is always equal to 1. If the hand is tilted all the way to the left along the X axis, the fundamental frequency is given a weighting of 1; if the hand is tilted all the way to the right, the high partials are given a weighting of 1; if the hand is tilted all the way forward along the Z axis, the middle partials receive all of the weighting. Everywhere in between, the X and Z tilt data is combined, and the timbre is determined by a weighted sum of all three stimuli [see fig.2].

X-tilt data from the "frequency" hand determines the value of the fundamental frequency, ranging from approximately 50Hz to 615Hz (roughly equivalent to the range of the human voice). The Z tilt controls the overall amplitude of the sound.

An emergent property of our application of the tristimulus timbre model is that it sounds like a variable formant synthesizer. Vowel sounds can be synthesized using a combination of three to four formants (discussed in the next section), a process very similar to tristimulus timbre shaping. Through careful control of the "timbre" hand, one can create a wide variety of vowel-like sounds, for which the human auditory system is innately tuned. In practice, a performer could learn to control the SoniMime system to create melodies with distinct vocal qualities.

3.2. Formant Synthesis as a Sonification Tool

The recognition and categorization of vowels is a highly developed feature of human auditory perception, and because of this, has enormous potential as an intuitive sonification tool [7]. The human ear is most sensitive to spectral peaks, known as formants, when identifying dissimilar sounds (such as human vowels). Formant frequencies differ based on speaker type (man, woman, child). Fundamental female formants are usually higher in frequency than those of males (with the exception of the 'ah' sound where the fundamental formant frequency for males is higher than that of a female), and formants of a child lie in a range between 60 to 180 Hz higher than those of a female.

In the current mapping scheme, movement data is used to control frequency, weighting of partials, and overall amplitude. The mapping scheme could be augmented to include control over other parameters, such as amplitude and bandwidth of formant peaks, further refining the qualities of auditory feedback.

4. RELATED WORK

Cassidy, Berger, and Lee use a weighting scheme for formant synthesis for the auditory display of hyperspectral colon tissue images [7]. In their experiments, they used a combination of four formant frequencies, distributed between the male 'i', the female 'a', and a child's 'u'.

Others have investigated speech synthesis using a combination of formant synthesis and other methods. In particular, Fels and Hinton in Glove-TalkII [8] explore speech synthesis using specific learned hand gestures in an adaptive interface. Also, Morrison et al [9] have developed a working prototype of a multiparametric dual-hand interface for real-time synthesis of expressive speech [10].

Instead of a static gesture-to-consonant mapping, SoniMime utilizes a truly continuous sensing environment that depends upon the user's fine motor control and ability to learn to manipulate a virtual timbre space, dictated by our tristimulus timbre model implementation.

5. FUTURE WORK

At the time of this writing, SoniMime is limited in versatility due to the hardwired serial connection to a computer. The authors are currently collaborating with Adam Bowen [11] to implement a version of SoniMime using Bluetooth for wireless communication between two small handheld devices and a macintosh powerbook running Max/MSP [see fig. 3].

The sonification of movement can be a powerful learning tool for physical activities as well as for data exploration. Sonification introduces auditory feedback in situations in which audible stimuli would otherwise be absent. Other future projects include expanding the spectrum of useful applications for SoniMime to include the sonification of specific learned motions, perhaps as diverse as choreographed dance sequences and even a golf swing.

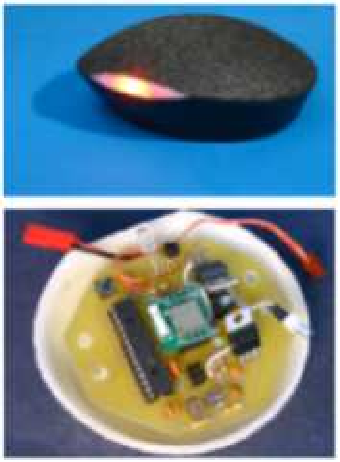

Figure 3: Wireless Handheld and Interior Circuitry

6. ACKNOWLEDGMENTS

SoniMime is the result of a project developed during the Autumn quarter in 2004 at CCRMA as part of a course taught by Jonathan Berger called "Auditory Remapping of Bioinformatics". Many thanks to Jonathan, Bill Verplank, and Max Mathews. The authors would also like to thank Matt Wright for the use of his accelerometer data filtering scheme, and Wendy Ju for her invaluable advice during the development process.

7. REFERENCES

[1] H. Pollard and E. Jansson, "A tristimulus method for the specification of musical timbre," *Acustica*, vol. 51, pp. 162 – 171, 1982.

[2] A. Riley and D. Howard, "Real-time tristimulus timbre synthesizer," 2004, http://www-users.york.ac.uk/ dmh8/ tristimulus.htm.

[3] M. Fernström and C. McNamara, "After direct manipulation - direct sonification," in *ICAD98*. 2004, British Computer Society.

[4] "Procyon engineering," http://www.procyonengineering.com.

[5] D. Levitin, M. Mathews, and K. McClean, "The perception of cross-modal simultaneity," in *International Journal of Computing Anticipatory Systems*. 1999, Chaos.

[6] "Sonimime," http://ccrma.stanford.edu/ jcarlile/250a/ sonimime.html.

[7] Ryan Cassidy, Jonathan Berger, and Kyogu Lee, "Auditory display of hyperspectral colon tissue images using vocal synthesis models," in *ICAD04*. 2004, International Community for Auditory Display.

[8] S. Fels and G. Hinton, "Glove talk ii: A neural network interface which maps gestures to parallel formant speech synthesizer controls," *IEEE Transactions on Neural Networks*, vol. 9, no. 1, pp. 205 – 212, 1998.

[9] G. Morrison, A. Hunt, and J. Worsdall, "A real-time interface for a formant speech synthesizer," *Logopedics Phoniatrics Vocology*, vol. 25, pp. 169–175, 2000.

[10] A. Hunt and T. Hermann, "The importance of interaction in sonification," in *ICAD04*. 2004, International Community for Auditory Display.

[11] "Adam bowen's portfolio," http://groove.morpheus.net/ portfolio/portindex.html.