The Sonification and Learning of Human Motion

CSUCI Research Competition Summary

Kevin M. Smith

February 12, 2014

Abstract

This paper examines how sonification can be used to help a student emulate the complex motion of a teacher interactively with increasing accuracy and timing. The system captures the motion of a single joint (such as a hand motion or gesture) of a *teacher* in real-time. A 3-D motion path is then generated and recorded along with a reference sound. A *student* then attempts to perform the motion and thus recreate the teacher's reference sound. The student's synthesized sound will dynamically approach the teacher's sound as the student's movement becomes more accurate. Several types of sound mappings which simultaneously represent time and space deviations are explored. For the experimental platform, a novel system that uses low-cost camera-based motion capture hardware and open source software has been developed. This work can be applied to diverse areas such as rehabilitation and physiotherapy, performance arts and aiding the visually impaired.

1 Introduction

In this paper we explore the use of sonification as a feedback mechanism for learning complex motion in real-time. Most techniques for learning movement, whether they are used for dance, sports or other articulated motion are visual. These can include interactive methods such as live demonstration and/or video recording of the movement for playback and coaching. These methods are predominantly visual with assisted verbal instruction provided by a teacher or coach. Here we focus on a different approach – the augmentation of the learning process with layered sound to provide a sonic feedback mechanism for learning precise motion.[1]

With the advent of relatively new and inexpensive technologies such as the Microsoft Kinect (2010)[1] which is enjoyed by many in the consumer games market, it is now possible to capture elements of human motion in 3-D and in real-time on relatively low-cost hardware such as a laptop computer. Inspired by these new developments, we specifically explore the idea of using sound as a feedback mechanism for learning how to reproduce a 3-D motion path generated by a joint on the body. In particular, we are interested in looking at how *layered* sound with varying pitch and loudness can be used to provide concurrent feedback on both spatial accuracy and timing of the motion. While we focus on a single joint or limb, we aim towards extending these principles to an entire body.

There is an abundance of existing work in the area of aiding movement by the use of sound. Rober and Masuch[2] used interactive auditory environments and 3-D sound rendering to explore virtual environments. Effenberg[3] used sonification to assist in the reproduction of human movements, showing that sound can provide additional information in the accurate reproduction of jumps and other athletic movements. Other work includes the use of sound for physiotherapy. Feedback is an area studied by Pauletto and Hunt[4]. In their work, the sonification of EMG signals gathered in a clinical environment

[1] A short video demonstration of the prototype system can be viewed at: http://www.youtube.com/user/k2msmith

Figure 1: Captured motion is reproduced using our sonification as a learning aid with our system, SoundTracer

provide auditory display to the therapist in real-time, PhysioSonic[6] was developed as a system to map motion capture data to sound to provide auditory feedback for physiotherapy and training.

Building on this existing work we specifically look at the use of sonification feedback for learning precise motion along a path with spatial and timing accuracy with focus on the development of a new portable laptop system which we call Sound*Tracer*. Figures 1 and 2 show examples of of Sound*Tracer* in action. The motion of any joint or limb can be captured in real-time for later retrieval. Sound is synthesized with the motion for auditory feedback during playback and when learning.

In the first section of this paper, we describe the overall system design including the sound mapping process. This will be followed by a description of the actual implementation of our experimental system and prototype. The implementation section will look at some of the initial results of using our system. We will then present conclusions and opportunities for future work.

2 System Overview

In our workflow, there are two primary actors, the teacher and the student. The *teacher* creates a movement that is captured and recorded in 3-D space in real-time. The *student* then has the goal of learning the motion created by the teacher by performing the motion as accurately as possibly. A successful performance is determined by the ability of the student to reproduce the motion of the teacher as accurately as possible with the same trajectory and correct timing.

The principal goal of the system is to use sound as a feedback mechanism to assist the student in learning the motion. Please refer to Figure 3 for a description of the process.

We capture the teacher's motion in real-time using the Kinect device (1). An internal representation in the form of a 3-D motion path is stored (2). Sound is generated to accompany the motion and the path and the sound are stored in a track (3), which can be recalled for playback (4).

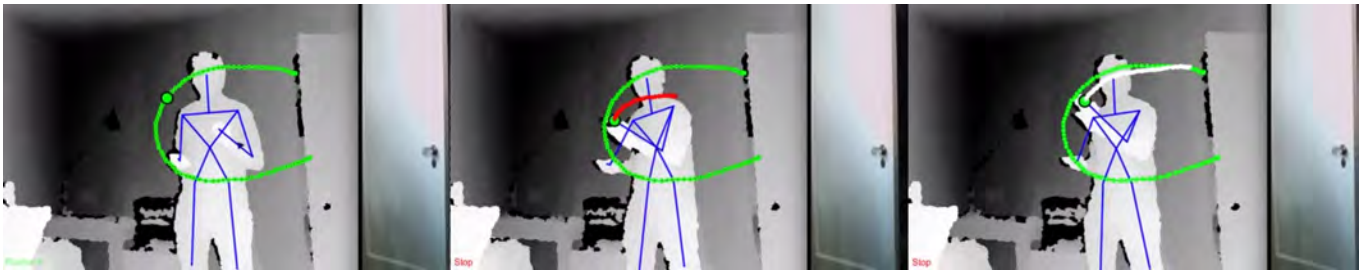

Figure 2: Student attempting to reproduce motion path. Student's path turns white as it approaches the reference path.

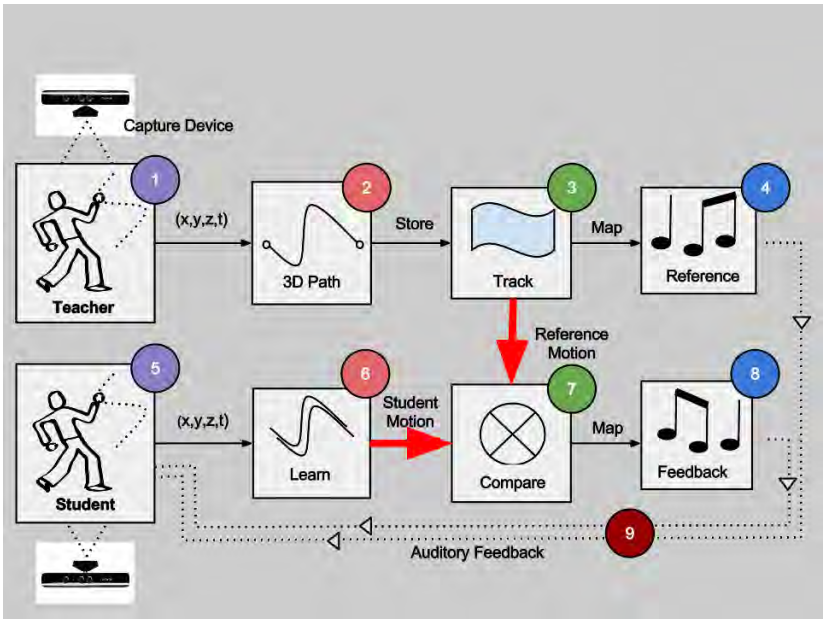

Figure 3: System Workflow: difference comparison between sonification of the teacher's motion and the student's.

Figure 4: Deviations from Teacher's path influence "pitch bend" parameter in sound mapping.

At the learning session (which can be done at another time), the students motion is captured in real-time using the Kinect device (5). The students motion is similarly converted to a track (6). While the student is moving, sound is generated by the student and compared in real-time with the motion the teacher previously generated in the reference track (7). Trajectory and timing differences between the student's motion and the teacher's are used to modify the sound. The new sound is provided as feedback to the student to aid in making corrections to the motion (9).

2.1 Sonification Process

The goal is to provide a transformed or "warped" sound to the student that approaches the teacher's sound as the student's motion gets closer to the teachers. Both spacial and timing accuracy can be tracked in real-time. The motion path is stored at each time interval in the form:

$$P(t) = (x, y, z, t)$$

Each *(x,y,z)* coordinate is captured in real-time at frame rate of 30hz[2] and saved in the track along with a time stamp *t*. The curve is stored as a piecewise linear representation and approximations of the curve length between two points or times on the curve can be quickly evaluated.

For spatial accuracy, we check the Euclidean distance in real-time with the closest point on the teachers path. Any differences in the distance can be used as an input to the sound mapping. In the figure 4, (from our prototype), the spatial error (shown in red) is used generate a "pitch bend" to vary the pitch of the teacher's sound being generated. When the motion is accurate, the pitch bend value is effectively 0 and the student's sound will be the same as the teacher's.

For temporal accuracy, we compare the student's current time and position with the on-target position of the teacher's at the same time.If the student is ahead of the on-target position (Figure 5), the time and distance error is provided as an input to the sound mapping. (i.e. how far ahead).

Similarly, if the student is going too slow and is tracking behind the teacher's target point, error parameters are

[2]There is no theoretical limitation. The frame rate is limited by the Kinect hardware and performance of the computer hardware.

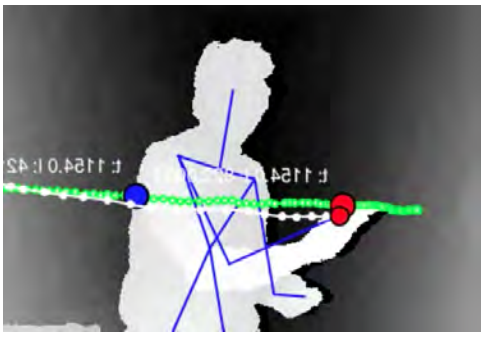

Figure 5: Student is going too fast and ahead of teacher's target point (blue marker).

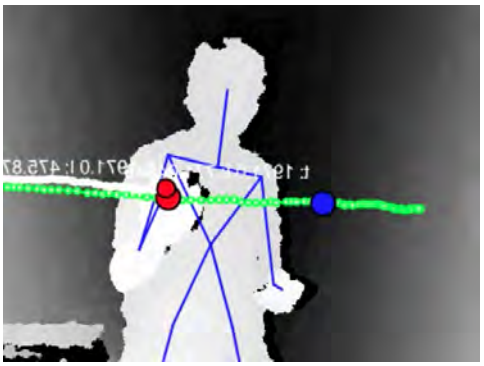

Figure 6: Student is going to slow and behind teacher's target point (blue marker).

provided to the sound mapping.

For the sound mapping, we developed a novel approach to layering an addition component to the sound to provide feedback on the timing. Taken from the sound synthesis field[10], we use an *attack, decay, sustain, release* (ADSR) envelope to control the volume of the sound output as the motion progresses along the curve. For our initial tests, we fond that a bell curve worked well although any envelope curve could be used in our implementation.

3 Implementation

We have developed a prototype system called Sound*Tracer* to enable a student to practice real-time reproduction of motion paths created by a teacher. A block diagram of the system is shown in Figure 7.

SoundTracer is based on open-source software components. We used Processing[8] as the implementation language because of its strength as a rapid prototyping language and the out-of-the-box experience enabling us to get it installed and running quickly with good library support for communicating with other devices and servers.

Although the code is generic enough to accept data from any 3-D tracking device, we used the first generation Microsoft Kinect device along with public domain libraries for the OpenNI[9] driver. The Kinect is limited in resolution and tracking stability in comparison to a professional optical motion capture system, however, the cost and convenience of this device made it practical for testing in most environments.

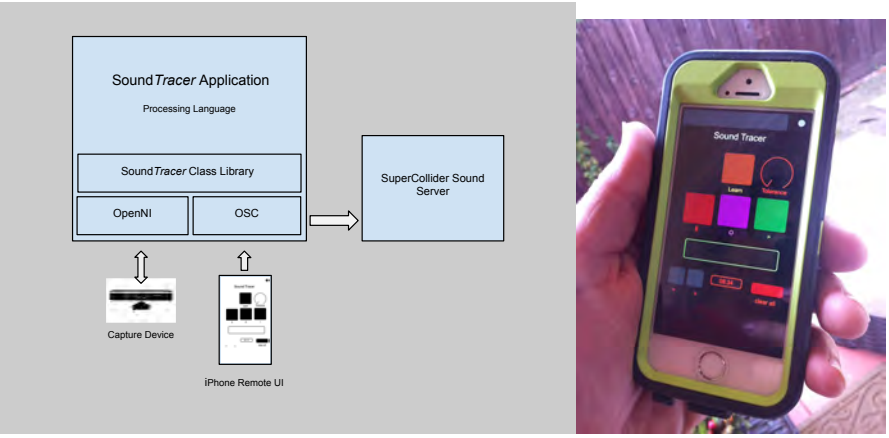

Figure 7: SoundTracer system consisting of Processing application communicating with Capture Device (Kinect), OSC control devices and the SuperCollider Sound Server.

The generation of synthesized sound for audio is a very mature field. Rather than implement our own software synthesizer, we chose to use the SuperCollider[10] language and server. For our reference sounds, a flute sound was generated using a waveguide flute example SynthDef in the SuperCollider language. The ADSR envelope was implemented in the same language using an envelope generator which can poll values by index (for more information see IEnvGen in the SuperCollider Reference)[11]

4 System Testing

To date we have conducted preliminary tests on the system with several users using the ADSR sonification method as described in Section 2.1. The goal of the testing is to provide initial feedback on the workflow process in order to increase usability of the system and to provide subjective feedback from the testers. Once this milestone has been achieved, we can form the basis for a more rigorous study in the future.

Three levels of sonification were tested: (1) Visual affordances only were presented to the user. No sonification of the 3-D data was used (2) Sonification + Visual. Sonification was used as an aid in addition to the visual affordances provided and (3) Sonification only. Sound was used as the main affordance. Minimal visual queues were provides. Note that in the last case, we still maintained some minimal visual queues to show the user where the motion path start/end is in order to provide a gate for measurement. In the case where visual affordances were used, we provided a display of the full motion path with target tracking markers on the path to show the student's current position with respect to the teacher's. The example data in Figure 8 shows a typical number of trials for a student for the Sonification + Visual case (2).

For each trial, three moves were scored; (1) horizontal planar motion, (2) vertical planar motion and (3) a more complex full 3-D motion path with movement axially in all directions. We used a root mean square technique to score the difference between the student's curve the teacher's reference curve (at each time interval). This calculation calculates both spatial and timing differences (points have time signatures). A lower number indicates a better score. Initial tests

User (n)	Trial (n)	Mapping (ADSR, SAMPLED)	Level (VIS, VIS+SON, SON)	Planar Horiz	Planar Vert	Mixed 3D
1	1	ADSR	VIS+SON	112	194	407
	2	ADSR	VIS+SON	68.7	55	227
	3	ADSR	VIS+SON	98.8	85	153
	4	ADSR	VIS+SON	82.7	47	109
	5	ADSR	VIS+SON	59.1	77	148
2	1	ADSR	VIS+SON	147	74.6	398
	2	ADSR	VIS+SON	81.2	48.9	294
	3	ADSR	VIS+SON	55.1	54.36	188
	4	ADSR	VIS+SON	77	76.46	146
	5	ADSR	VIS+SON	52.1	51.47	127

Figure 8: Example Sonification data for a test user (student).

indicate that with all three motion types (horizontal, vertical, mixed), we achieve a significant improvement in the score after five trials. We are currently doing additional focused testing with various combinations of sound only and visual.

4.1 User Feedback/Impressions

Following each test we collected feedback and the comments are summarized below:

- Visual Representation – The reference video on the screen (see figure 6) is rendered from the perspective of the camera, so the image is reversed from a "mirror". Most users preferred to see a mirror image of their body. This transformation has been implemented in the system.
- Ambidexterity – The initial system focuses on motion of a single part (joint) in the body. Most users were not equally as good at reproducing the teacher's path with both hands, particularly the spatial aspect.
- Tolerance – Reproducing a path given the accuracy of the hardware and the user required us to have a configurable tolerance, effectively converting the path from a line to a "tube". For coarse full-body movements, several centimeters might be acceptable, but for more fine-grain hand motion, smaller tolerances may be used.
- Sonification (Spatial) – Users all agree that the pitch change sonification helped them to know when and where they went off track, but it was not always clear which direction to move to correct. Currently we "bend" the pitch up linearly for deviations from the curve. We are investigating other mappings, which can use direction movement for pitch up/down. Sometimes the visual represents confused this further and better results were obtained from a user when the visual affordances were turned off.
- Sonification (Temporal) – Initial testing impressions of the ADSR envelope method were favorable. In particular, playback of the teacher's envelope prior to each exercise enabled them to develop a mental image of the sound to make when the correct timing is achieved.

- SCORING – Our initial attempt was to simplifying scoring including both spatial and timing performance in one number. It became quickly apparent that we need to break this up into multiple components, so that the student can understand what needs to be improved, either timing or trajectory.

5 Conclusion and Future Work

Our basic testing of Sound*Tracer* as an experimental platform for motion sonification and learning has yielded some initial results that are very encouraging. We have created a useful system that can map real-time movement to layers of sound, which can be used as a feedback loop for learning complex path-based motion. We are looking forward to further testing with more complex motion paths and sound mappings. In this initial effort we have focused on comprehensive path animation of a single point on the body over time. As a next step, we can explore how to scale this to a fully articulated hierarchical skeleton. This may involve capturing the hierarchical motion of each joint or limb or perhaps investigating combining these methods with a pose-based approach. At a smaller physical scale, capturing more detailed skeletal features such as the hand could allow us to capture finger/hand based motion paths. Our method could prove to be a powerful technique of augmenting a gesture-based vocabulary with motion path information that includes both timing and spatial information.

6 Documentation

A short video demonstration of the system can be viewed at: http://www.youtube.com/user/k2msmith

References

[1] http://www.xbox.com/en-US/kinect

[2] Niklas Rober and Maic Masuch, "Interacting with Sound, An Interaction Paradigm for Virtual Worlds", Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sidney Australia, July 2004.

[3] Alfred O. Effenberg, "Using Sonification to Enhance Perception and Reproduction Accuracy of Human Movement Patterns", Proceedings of the Int. Workshop on Interactive Sonification, Bielefeld, Germany, Jan 2004.

[4] Sanda Pauletto and Andy Hunt, "The Sonification of EMG Data", Proceedings of the 12th International Conference on Auditory Display (ICAD 2006), London, UK, June 2006.

[5] Thomas Hermann, Andy Hunt, John G. Neuhoff, The Sonification Handbook, Logos Verlag, Berlin, Germany 2011.

[6] Katherina Vogt, David Pirro, Ingo Kobenz, Robert Holdrich and Gerhard Eckel, "PhysioSonic – Movement Sonification as Auditory Feedback, Proceedings of the 15th International Conference on Auditory Display (ICAD 2009), Copenhagen, Denmark, May 2009.

[7] http://ecmc.rochester.edu/ecmc/docs/supercollider/scbook/

[8] http://www.processing.org

[9] http://www.openni.org

[10] Scott Wilson, David Cottle and Nick Collins (edited), The SuperCollider Book, The MIT Press, Cambridge, Mass, 2011.

[11] http://doc.sccode.org

[12] http://www.sojamo.de/libraries/oscP5/

[13] Adrian Freed, Andy Schmeder, "Features and Future of Open Sound Control version 1.1 for NIME", New Interfaces for Musical Expression (Conf) 2009, Pittsburgh, PA, June 2009.

[14] http://hexler.net/software/touchosc