Teaching Rhythm In Childhood Music Education Through Gesture Controlled Loop Generation

Jeremy Friedman

MASTER THESIS UPF / 2012
Master in Sound and Music Computing

Master Thesis Supervisor: Sergi Jordà

Department of Information and Communication Technology

Universitat Pompeu Fabra, Barcelona

UNIVERSITAT POMPEU FABRA

I dedicate this thesis to Amaia whose commitment to educating children is nothing short of inspiring. And to Ona, who is the greatest teacher I have ever had.

Acknowledgements

I would like to thank my SMC colleagues and professors for an amazing year. I have learned so much. I would especially like to thank Aluizio Barbosa for all of his help with this project and Sergi Jordà, whose guidance and insight consistently brought out the best in me.

Abstract

This thesis explores the viability of bringing gesture recognition technology to the classroom environment to aid in teaching rhythm using the Microsoft Kinect camera. The use of movement games to teach rhythm in childhood music pedagogy dates back to the early 20th century with well-established methods including Dalcroze Eurythmics and Orff Schulwerk but the overall structure of public education has changed very little since then, creating an incongruity between antiquated teaching methods and the world of rapidly changing technology that surrounds the children of the 21st century. The need to make learning current and relevant has inspired many independent initiatives to use the Kinect in the classroom but there is very little literature formally studying its implementation. This study looks at the usability, with regards to rhythmic complexity and modality, of a gesture-based teaching aid using the Kinect with children ages 8 – 10.

Keywords: Kinect, Processing, Pure Data, gesture recognition, childhood music pedagogy, Dalcroze, Orff

Table of Contents

List of Figures

1. Introduction

1.1 Motivation

The current public school teaching system was developed on values stemming from the cultures of the 18th and 19th centuries of enlightenment and the industrial revolution, giving root to the issues of how to make learning current, relevant, and fun to children of the 21st century who live in a world of mobile phones, touch tablets, and interactive gaming [1]. The problem is only exacerbated by the current economic policy that does not treat education as a worthy investment and forces learning institutions to do more with less [2][3][4], severely limiting the possibility of keeping up to date with the world of technology that surrounds our children.

The development of new software and interfaces helps to generate different paradigms and solutions in music education [5] and the accessibility of current technology has inspired several independent initiatives to incorporate open-source software and "hacks" in the classroom [6] providing creative low-cost educational alternatives that engage children to be in the moment and gets them excited about learning [7].

1.2 Research Goals

The main goal of this research is to develop a teaching tool that would aid in the teaching of rhythm in a music classroom environment. In designing a viable system it had to exhibit certain functional qualities:

- Effective: to be educationally meaningful it was important to base the system on established methods of music pedagogy and current cognitive learning theory.
- Engaging: the tool needed to be interactive and fun for students and teachers alike.
- Noninvasive: the interface design could not disturb the natural dynamic of the classroom and or in any way draw focus from learning.
- Economical: the system should be implemented using software and hardware that was easily accessible either free or inexpensive.

2. State of the Art

2.1 What is Rhythm?

Rhythm is something that is universally felt and experienced but is often hard to define. In their book *A Generative Theory of Tonal Music,* Lerdahl and Jackendoff describe rhythm as a cognitive structure that is made up of perceptual dimensions allowing us to "chunk" and process information. Within this main structure are two independent, yet interrelated, components of rhythmic organization - the grouping structure and the metrical structure. Groups are a hierarchical, non-overlapping, recursive, and contiguous organization of sound signals, i.e. motives, phrases, and sections - while meter is a regular pattern of strong and weak beats to which we relate the sound signal. Beats, in this context, refer to a pattern of accentuation at the musical surface called metrical accents that provide a mental construct inferred from said patterns. This grouping is directly related to a perception of a series of regularly occurring stimuli marked for consciousness that provide a contextual weight in a periodic pattern giving us a strong sense of grouping or pulse. Together, these dimensions create a complex mixture of meter and grouping constraints in reference to a beat which we call rhythm [8].

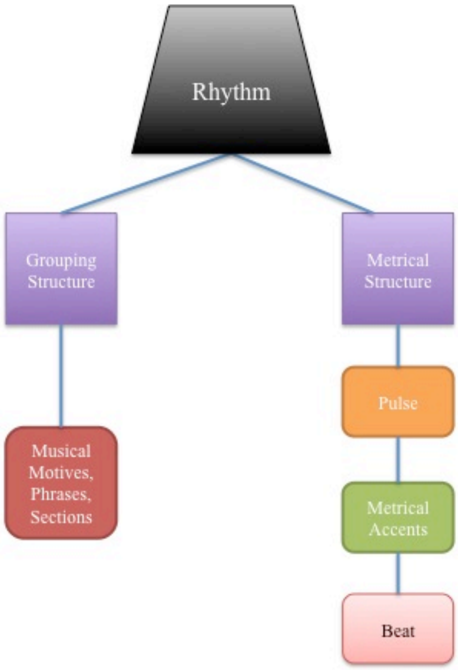


Figure 1: Visual representation of rhythm hierarchy

Research shows that our cognition of musical events is present from birth. In a study using a combination of EEG and motor response analysis in neonates, there is evidence that we can distinguish between organized music and random sounds from infancy [9]. Furthermore culturally independent rhythmic recognition is developed from as early as six months old and a clear cultural dependent rhythmic recognition develops around the age of one [10][11].

It is obvious that rhythm is a key element in our perception of music and its importance in childhood music education can not be overstated as it is in this very arena that foundational experiences are provided to reinforce the fundamentals of music [12].

2.2 Methods and Concepts in Music Pedagogy

There are a wide variety of teaching styles and pedagogical methods used in the instruction of abstract musical concepts. Traditionally these concepts are taught in relation to a musical instrument where each concept is compartmentalized and studied using exercises and etudes, with the goal of improved performance [13][14]. In teaching general music, often the first classes start with the concepts of volume and tempo, and are quickly followed by pitch. These concepts are taught roughly from the age of four. In later lessons, the concept of timbre is introduced. From about the age of eight, classes focus on the more difficult concepts such as tone, duration, rhythm, harmony and articulation [15].

The appearance of methods that use a more holistic approach to music instruction and employ the concept of embodiment to their teaching have greatly influenced the music education world [16].

a) Dalcroze Eurhythmics and Orff Schulwerk

In the early 20th century, Swiss music teacher Emile Jacques-Dalcroze (1865-1950) developed a system of rhythmic gymnastics or 'eurhythmics' to encourage a competency in enacting rhythms, particularly polyphonic rhythms [17]. Amongst other things, this involved asking students to walk at a regular pace, while moving their arms in synchrony at, for example, twice or three times the rate. Dalcroze's approach and its modern refinements represent established trends within rhythm education [18].

To effectively communicate to children, the Dalcroze method builds on a rhythmic vocabulary starting with rhythmic patterns that correspond to the beat that generally hold up to four elements, and are presented gradually, starting from the most regular (binary). This vocabulary is experienced by body movement, body percussion, and through the voice [19]. The Dalcroze method inspired other methods including Kodaly (1935) and Orff Schulwerk (1924) where the key to learning is based on the way children play through imitation, exploration, improvisation, and composition [20].

b) Rhythmic Embodiment

Embodiment can be defined as the phenomenological paradigm that emphasizes the role of action and perception in meaning making. Studies in embodied music cognition explain that the body is a natural mediator between the mind and the physical environment, and, as opposed to the disembodied perception-based analysis of music structures, meaning is given to music through corporeal and kinesthetic means [21]. With respect to interaction, this puts the emphasis for the understanding and incorporation of our relationship with the world around us, both physical and social, into the design and use of interactive systems [22].

Experiments in auditory coding of rhythmic movement show that how we move will influence what we hear. In a study to provide an empirical basis for Dalcroze's Eurhythmics, a series of experiments showed that cognition results from a combination of physical experience and strong internalization of metrical structure requiring the active involvement of the body, which later aids in identifying ambiguous rhythm [23].

Going a step further it has been shown that specific kinds of embodiment may be more effective using what is referred to as embodied interactional metaphors. Testing was done on children aged seven to ten in an interactive musical sound-making environment called The Sound Maker where the children manipulated the volume, tempo, and pitch of musical sounds through body movement. Their natural gravitation towards spatial rather than body-based movements inspired the researchers to investigate various mappings in relation to embodied and ontological (the representation of an abstract concept with something concrete and physical) metaphors [15]. These metaphors were derived from image schemas, which are abstract representations of recurring dynamic patterns of bodily interactions that structure the way we understand the world.

Image schemas are schematic in nature as they capture the structural contours of sensory-motor experience and are not just symbols. They exist beneath conscious awareness and integrate information from multiple modalities represented visually, haptically, kinesthetically or acoustically. Although image schemas describe human experiences with the physical world their actual strength lies in their metaphorical extension for structuring abstract concepts.

The various schemas are broken into sub-categories: space schemas (i.e., concept of up-down), containment schema (i.e., physical or metaphorical boundary), force schemas, and attribute schemas (i.e., heavy-light, dark - bright) [24].

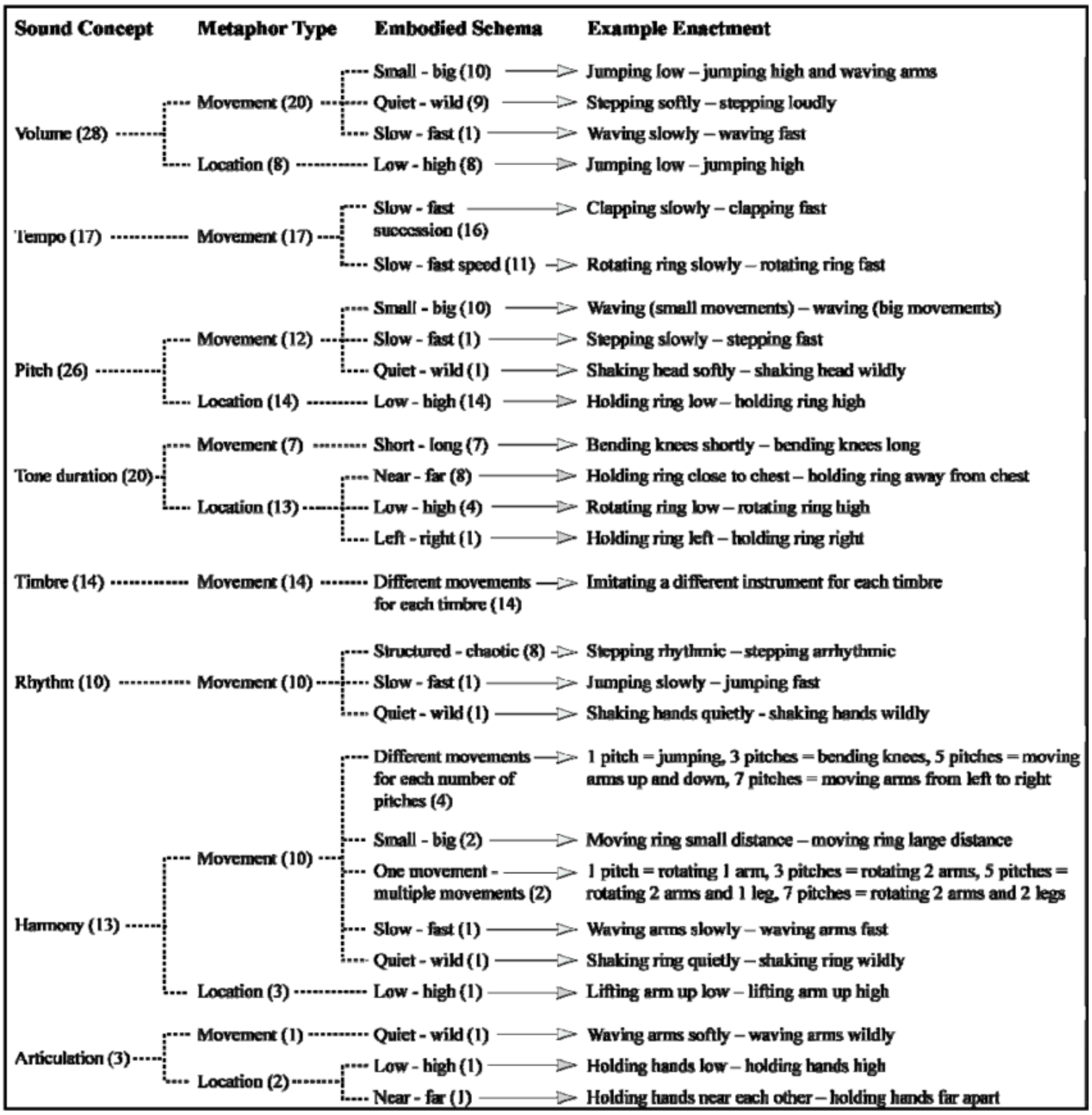


Figure 2: Chart attributing image schema and embodied metaphors to musical concepts.

Results of the Sound Maker experiment showed that children using the metaphors performed better than those not using the metaphors [23].

Later studies conflict on systems that incorporate multiple embodied metaphors, as opposed to one metaphor per sound concept. Some research shows that multiple metaphors help children identify with individual creativity, hence supporting conceptual learning in abstract domains [25] while other studies show that, while the use of multiple embodied metaphors do not confuse children, multiple movements do not necessarily improve their understanding [18].

c) Multimodality

An important aspect of music embodiment is the effect of sensory modes, specifically multimodality, on the learning process. There are conflicting studies on the importance of individual modalities but what seems true is that associations between modalities are formed at the beginning our lives. Experiments testing the ability of 7-month-old children to interpret strong and weak beats showed an inherent multisensory connection between body movement and auditory rhythm processing. This was concluded to be due to the early development of the vestibular system when, for example, a child has been rocked or bounced to a lullaby, creating a strong vestibular-auditory connection critical for human musical development [22].

With respect to the pedagogical weight of individual modalities, in a study of appropriate modalities for music learning applications, students between the ages of 19 and 24, of varying musical skill levels, and representing a wide range of instruments, were asked to rate modalities from most important to least important, with respect to their individual instruments. In all cases haptic clearly came out on top with the visual consistently having the least amount of importance [26]. Other research has shown that the visual modality, while not necessary for encoding rhythm, can enhance perception [22].

This idea was further explored in the design of a framework for an interactive sonification system applied as a practice tool for instrumentalists and singers. Removing the visual modality altogether, this system utilized auditory stimuli to foster an autodidactic approach to music pedagogy. By developing different auditory feedback cues that respond dynamically to a given performance, the use of interactive visualization is forsaken for interactive sonification [27]. It would appear from the given research that the effect of certain modes in music education is a subject that requires further investigation.

2.3 Technology in Music Education

Current research in incorporating technology into music education has taken several approaches using a variety of media including mobile devices, tangible artifacts, motion tracking, and social networking. In the MoBoogie project, an android application was developed with the aim of inspiring purposeful and emotion-evoking creative expression in children. Using the three axes of a mobile device's accelerometer, spontaneous loops were generated through movement and dance [28]. Another project, MOGCLASS, took advantage of the sound synthesis technology and sensory capabilities in mobile devices to simulate playing a wide range of musical instruments through body movements. The devices were networked to allow for group collaboration with the aim of enhancing active listening, composition, and performance in the classroom [29].

Group collaboration has also been researched with experiments using a web-based ensemble training application. With the aim of creating a tool for remote learning among classes and schools, rhythmic accuracy data is collected using accelerometer information from Nintendo Wii controllers in conjunction with a virtual drum set implemented via a graphic user interface on a computer screen [30].

Inspiring creativity not only in instrument performance but also construction, the CoolMag project investigated a system using finger magnets, magnetic sensors, and an Arduino SCM for a low cost solution towards instrument construction in the classrooms of developing countries. Students could build instruments from available materials, i.e. a "broom-guitar" or a "table-piano", and trigger sounds of that instrument from their hand movements via MIDI [31]. Another example that used movement tracking is the Harmony Space interactive learning environment where a user's motions are tracked inside of a boardgame-like interface to teach chord sequences through whole body navigation with the goal of providing rich physical cues for memory and engagement [32].

In making direct use of video games in music pedagogy, research in using the game Rock Band to teach drum set to students with no previous drumming experience showed a relatively high level of success when the subjects were transferred to a real drum set. Experiments also showed the importance providing visual and aural feedback within the video-game-as-didactic-tool environment [33].

2.4 Similar Work

Apart from the examples listed above, I would like to highlight three projects that are most closely related to the research presented in this thesis.

a) Wireless Gesture Follower

Developed inside of the i-Maestro framework [34], the goal of this research was to create a system that compared, in real-time, performed gestures with pre-recorded examples using machine-learning techniques. The system was made to help teach children smoothness and fluidity in conducting, which is an essential skill in music training closely related to music theory and sight-singing. To use the system, the teacher starts the exercise by conducting to a sound file, recording a reference gesture of their performance. The student continues the exercise by conducting the same sound file whose playback speed varies depending on the temporal variation of the gesture performance.

Figure 3: The wireless gesture follower.

Wireless sensor devices were attached to the users for motion tracking and data was sent via OSC to the host computer running Max/MSP [35]. The results proved to be encouraging regarding system robustness and approach but there was a clear need to conduct larger scale experiments to determine general viability in music pedagogy [36].

b) Sound Maker and MoSo Tangibles

The Sound Maker was a test bed to look for evidence that, by leveraging embodied knowledge, interactive environments could be designed to support children's learning of abstract musical concepts. Testing was performed with 20 pairs of children aged 7 - 10 who were asked to perform a series of musical tasks involving recognizing, mimicking and creating simple patterns in an interactive audio environment, relating locations, quantities and qualities of bodily movement to changes in the sound parameters of percussive audio output. Movements were recorded with video and data was collected through a color-digitized video capturing patch in Max/MSP/Jitter.

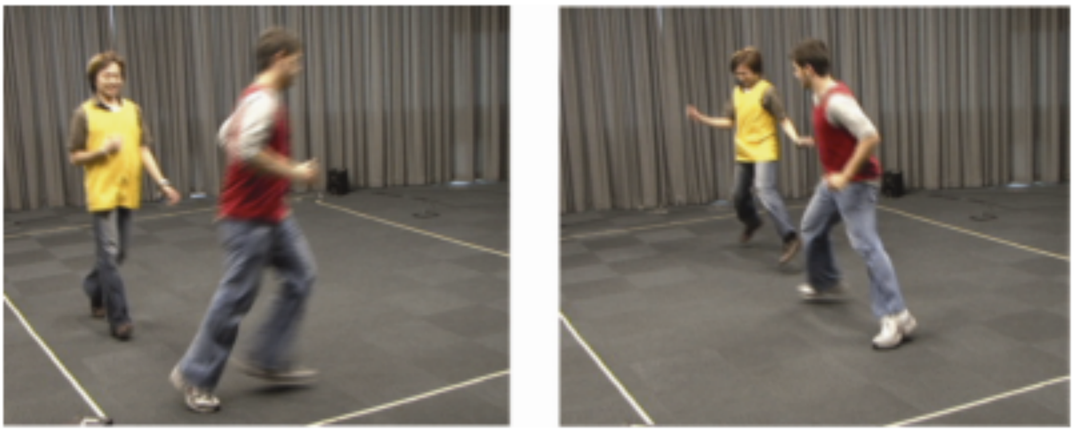


Figure 4: The Sound Maker

Results showed that children had much greater success in performing musical tasks when learning involved the use of embodied metaphors [37][23].

The Sound Maker project was later extended to explore if, and why, interactive systems that incorporate multiple embodied metaphors in their interaction models can support children's conceptual learning in abstract domains. The experiment was carried out using a specifically designed tangible learning system called MoSo Tangibles - a set of dedicated tangible artifacts that implemented three different embodied metaphor-based interaction models for learning about single sound concepts.

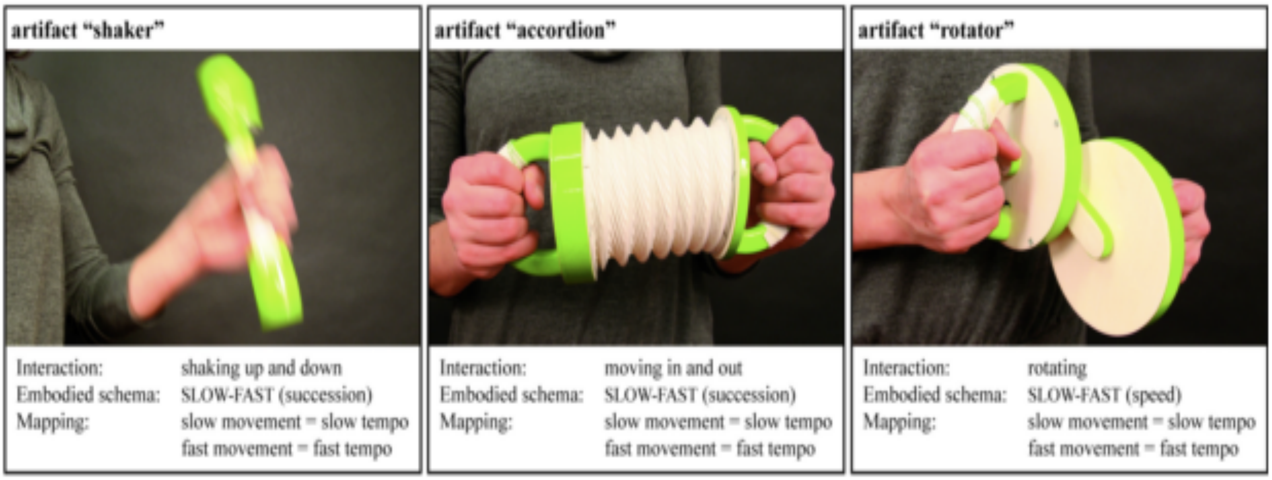


Figure 5: MoSo Tangibles artefacts and related schema.

While results showed no performance difference between multiple metaphorical mappings and single mappings, there was clear evidence supporting the use of embodied metaphors to teach musical concepts [24].

c) Haptic Drum Kit

The Haptic Drum Kit takes an original approach in attempting to recognize, identify, retain, analyze and reproduce rhythmic patterns on the drum set using a purely haptic learning environment. The system attempts to take advantage of rhythmic embodiment principles outlined by Dalcroze, as well as two other concepts: (1) sensory motor contingency theory, which sates that in order to learn to organize and respond appropriately to sensory input in some new domain or context, it is that the individual learner's motor actions have the power to affect relationships in the domain being sensed causing effects sensed by the individual's own sensory apparatus, and (2) entrainment, which describes how two or more physically connected rhythmic processes interact with each other in such a way that they adjust towards one another and eventually lock in to a common periodicity.

Vibrotactiles were attached to the ankles and wrists of the test subjects, who made up of both novice and experienced drum set players. The vibrotactiles delivered stimuli that corresponded to a rhythmic sequence that was written in Max/MSP and sent control signals was sent using an Arduino [38].

The test group was asked to perform rhythms taught to them either aurally or haptically. Tests showed that rhythms could be successfully learned haptically but the best results came about from a combination of the haptic and aural modalities [25].

Figure 6: Haptic drum kit.

3. The System

3.1 System Overview

The design concept for the system developed for this thesis focused on providing a teaching aid for music professors to teach rhythm to children in a music classroom environment. In the scope of achieving my first two goals of making the system effective and engaging, I drew from the strong evidence presented in research, and the established teaching methodologies, from the previous section, which supported the use of embodiment as an effective medium through which children can incorporate abstract musical concepts. It was shown that, through movement and rhythmic embodiment, children feel engaged in a learning activity that is reminiscent of how they play, and that creating an environment akin to a video game should enhance this sense of engagement by making the method of information transmission current and relevant with respect to the technology-filled world in which our children live today.

While the similar works outlined in the previous section made use of embodiment for teaching rhythm in childhood music education, specifically using ideas taken from the Dalcroze method, it seemed engagement could be further enhanced through creative input on the composition of the rhythmic tasks, leading to a task design for my experiments that is more of a structured outline with room for improvisation, as opposed to a strict replication.

Furthermore, to maintain the learning dynamic, I wanted my system to be noninvasive. Attaching physical sensors on children was out of the question in terms of setup time and the amount of material needed for an entire class. It was also important that at no point the children felt constrained by having to wear a physical apparatus. As well, the use of tangibles, while possibly enhancing embodied metaphors and image schema, involves the construction of specific materials and requires explanation regarding their use. It seemed important for the class flow that the children made use of a skill set that they already possessed and did not have to delve into new instruments for learning.

Lastly, the previously presented systems used relatively expensive specialized software and/or hardware, and to fulfill the requirements of a system design that made use of technology that was accessible and economical, my research used only open-source software and a Microsoft Kinect camera, which has an incredible accessibility factor due to it's popularity as a video console peripheral.

In the end, a system was developed allowing users to create spontaneous loops by triggering sounds using previously learned gestures. Working inside the framework of a game, classroom colleagues would be challenged to replicate the loop in an attempt to obtain the highest score. The parameters of the game would be controlled by the teacher of the class using a mobile device with a custom GUI, as to allow for the operation of the system regardless of level of computer aptitude, and the individual scores would provide an empirical method to gauge rhythmic timing accuracy. A minimalist visual representation that corresponded to the generated loop was also implemented to provide visual feedback for the users, which, for the experiment, could be switched on and off, to determine how it affected the performance scores.

In the following sections I will outline the individual conceptual elements and the technology used in each step of the design.

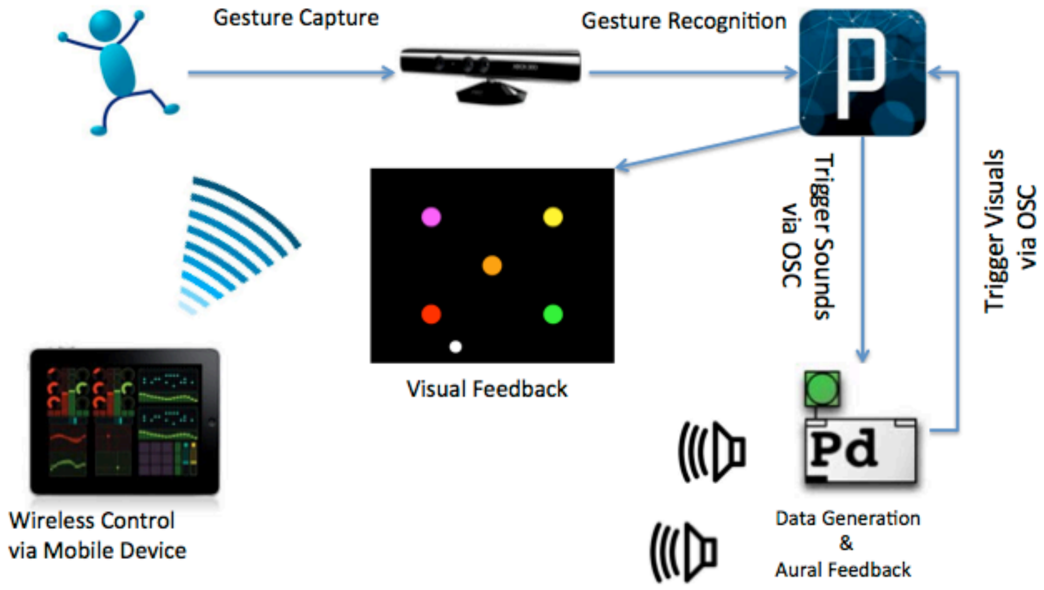


Figure 7: System overview.

3.2 User Detection and Motion Capture

Motion capturing was performed using depth images taken with the Microsoft Kinect camera [39].

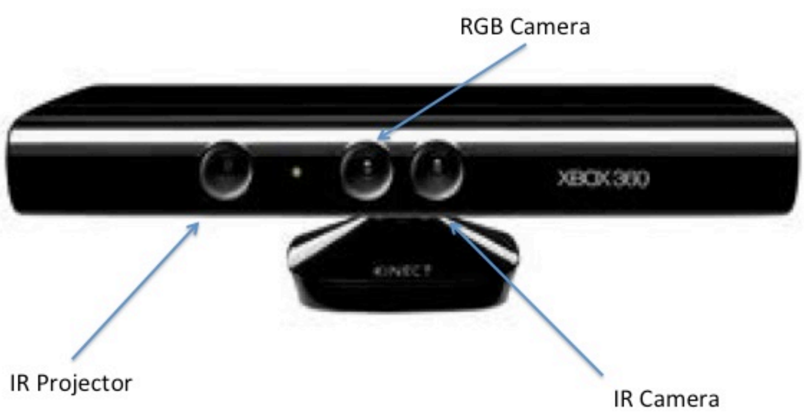


Figure 8: The Kinect

The Kinect is marketed my Microsoft as a peripheral for their Xbox 360 video game console allowing users to control games with gestures. The device houses two cameras - an RGB and an infrared camera - and an infrared projector that enables it to take depth images at 30 frames per second [40]. The Kinect offers several advantages as a motion capture device in an educational scenario including familiarity with interaction from its use with video games, low cost, a large community developing hacks for the Kinect expressly for educational purposes, and the ability to access 3D positioning [7]. The Kinect has the potential to enhance classroom interaction and student participation due to its support of multiple types of interaction, to create enjoyable and interesting experiences and boosts motivation. Constraints include a lack of standard development and long calibration time [41].

The image data was accessed implementing Simple Openni [42] in the open-source software Processing [43]. Simple Openni is a wrapper of the Prime Sense software OpenNi and the middleware Nite to provide the functionality of these two libraries in Processing [44].

The first step in user identification uses a technique called Center of Mass, which passively detects users by calculating the center most position of their mass [45]. Upon calibration, skeleton tracking begins and joint positions can be accessed and used to define and recognize gestures.

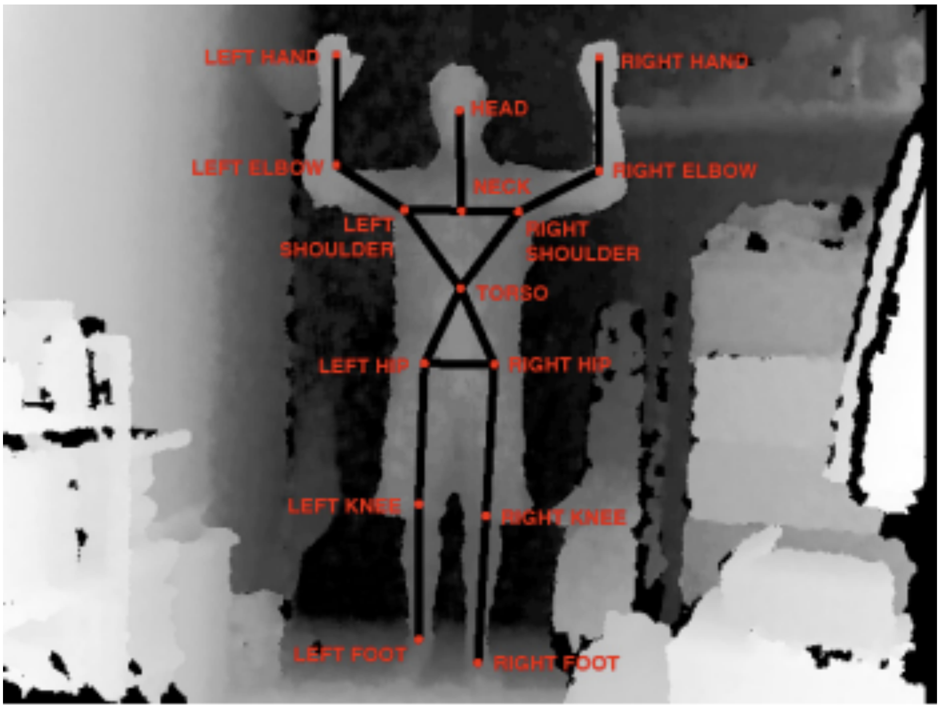


Figure 9: Depth image of calibrated skeleton and joint positions taken with the Kinect.

3.3 Gesture Recognition

a) Developing a gesture library

In deciding on the gestures that were to be recognized I wanted to draw from movements that were already being used by the testing group with the aim of developing a gesture library that was meaningful and relevant. It was also important to optimize the ease of interaction by not introducing new gestures that would have to be learned [46]. The music classes of the test group were taught using the Orff Schulwerk method where rhythmic embodiment is a core element of the music lessons [20]. After observing classes and working with the music professors on a reasonable number of gestures we decided on head pats, lap slaps, and clapping, for a total of five gestures, taking into account both limbs.

b) Gesture Definition

Upon successful joint skeleton tracking, the gestures were defined using vector subtraction to measure the relationship between two joints in depth pixels:

$$G = \vec{J_1} - \vec{J_2}$$

$$|v|_{(G)} = \sqrt{x^2 + y^2 +} z^2$$

$$\hat{v} = {}^{|v|}/_{||v||}$$

Where G is the individual gesture and $\vec{J}$ is a joint vector, the unit vector $\hat{v}$ is derived by dividing the vector magnitude $|v|$ by its norm.

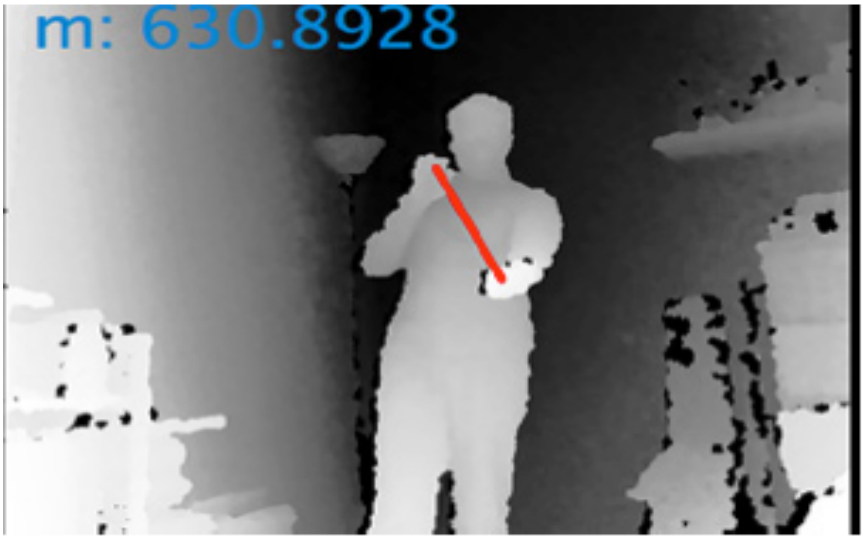


Figure 10: Vector subtraction between the hands with distance shown in pixels.

The subtracted vector for each gesture was sent to Pure Data via the OSC protocol and testing was done to determine an approximate pixel threshold below which a gesture could be considered recognized (~250 pixels). The thresholds were designed to be independently variable in real time to account for the varying joint distances of different users, i.e. between an adult and a child. A variable smoothing - between 10 and 20 milliseconds - was also applied to reduce noise and prevent the double triggering of the sound samples.

3.4 Aural Feedback and Latency

The recognized gestures triggered pre-recorded samples in Pure Data to provide aural feedback to the user. Although the study was based on rhythm it seemed apt to use tonal percussive sounds that could take advantage of pitch for the indication of zones in the body, as well as to distinguish between left and right using low and high sounds (like a keyboard).

It was also important to take into consideration a user's perceived timing latency from when a gesture would be performed to when feedback would be received. As measuring rhythmic accuracy is the main focus of the system, I needed to keep the latency within a range that has been shown to be acceptable in most musical situations (20 - 30 msec) [47]. To determine the amount of latency I performed several trials with the system, using simple rhythms, and measuring the time between a performed gesture and the perceived feedback. As a conservatory trained percussionist with over 12 years of professional experience my perception of latency would most likely be more refined than the children in the test group [48].

Taking the average of the results of several trials the perception of zero latency was produced delaying the audio feedback by the resulting value (~ 94ms). This is discussed in more detail in section 4.2.

3.5 Visual Feedback

The importance of the visual modality in teaching and learning rhythm was discussed in section 2.2.c and examined in research presented in sections 2.3 and 2.4. A survey of modalities rated visual feedback as the least important modality [26] yet research has shown that it can enhance performance [22] and is some cases essential to the performance [33]. As mentioned in 3.1, I wanted to measure the importance of the visual modality to the system by interchanging the provision of visual feedback with a lack of feedback, and observe how that affected the results of in terms of rhythmic accuracy.

To provide the user with visual feedback, the recorded gestures were sent back to Processing, via the OSC protocol, to a dynamically generated feedback score, which featured colored indicators displaying the position of the gesture within the score, along with a metronome marker indicating the position within the measure and beats of the score. It was also decided that there needed to be a visual pre-indicator that would signal the upcoming timing position and gesture of a rhythmic event to make the system more playable, as there would be no rehearsal before loop replication.

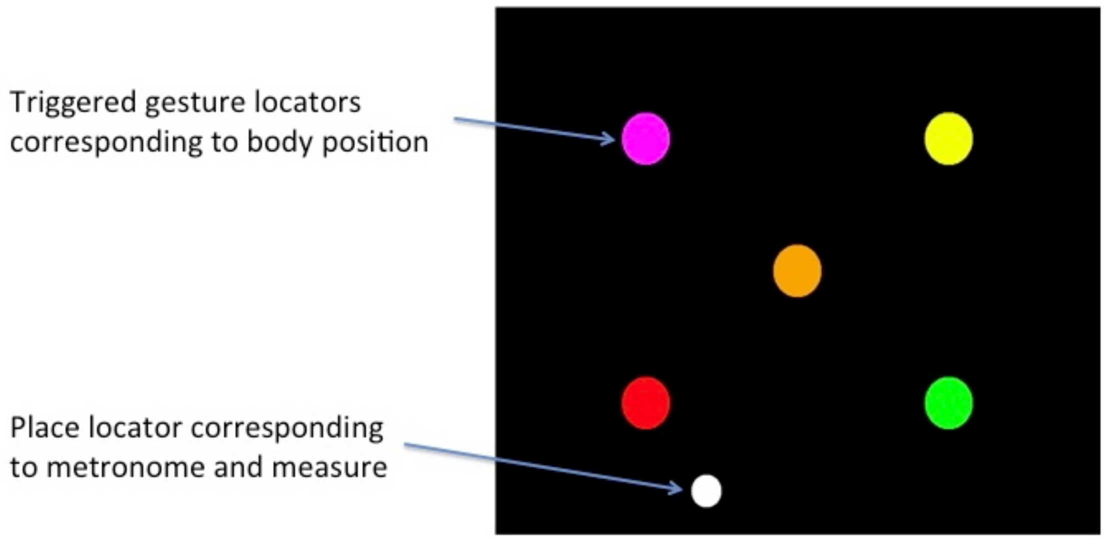


Figure 11: Visual feedback interface.

3.6 Control Interface

As was discussed in section 3.1, a custom controller GUI on a mobile device was designed to give the professors of the test group the ability to control parameters of the experiment regardless of computer skill level. The control interface was designed in TouchOSC [49] and implemented on an Apple iPad 2, offering full control over the necessary elements of the system including starting and stopping an individual test, tempo adjustment, independent volume sliders for the metronome click and the samples, and a text file generator for the results of a test. There were also feedback indicators including dynamic score readouts for each gesture, the total score, and an indicator of time position within a given iteration of the experiment.

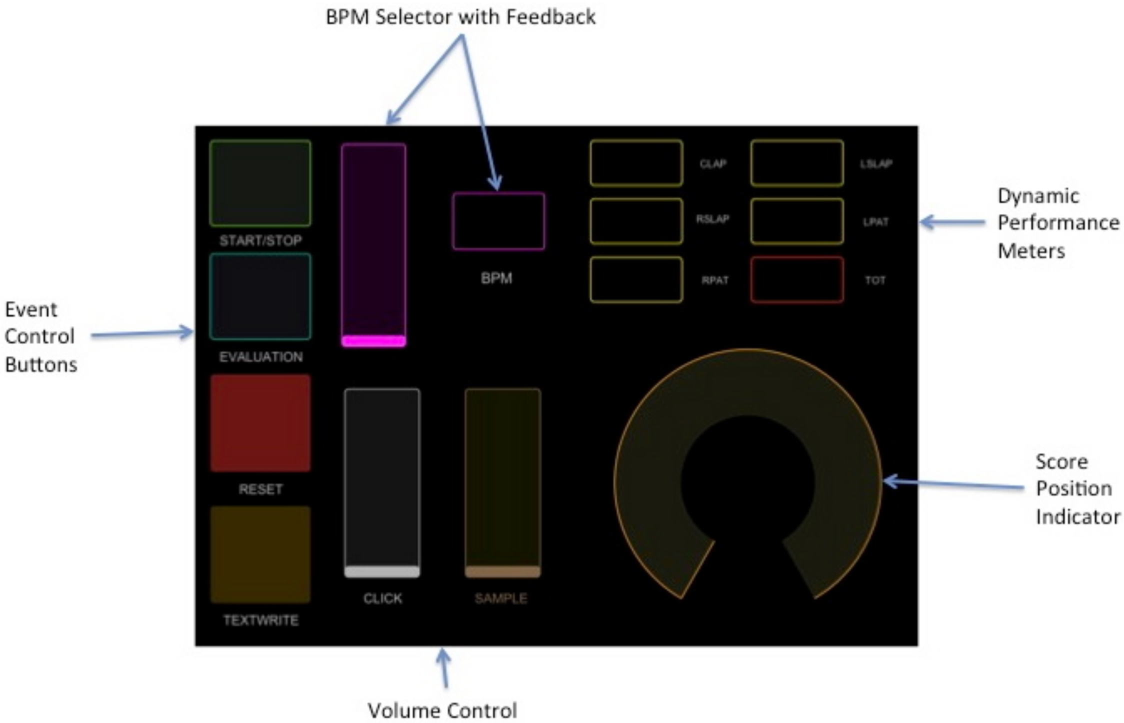


Figure 12: Custom GUI using TouchOSC.

4. Methodology & Evaluation

4.1 Testing Method

The main focus of the experiment was to evaluate the ability of the system to test rhythmic timing accuracy, as well as to measure the effect of the visual modality on the accuracy. The test was performed twice - at the Escola Municipal de Música de Castellbisbal with a group of 10 children, boys and girls, between the ages of eight and ten, where the children had been studying music between four and five years and all played an instrument, and at a private home with four boys between the ages of seven and eight, with varying levels of music study.

Random pairs were formed from the test group and the system was explained to them as a game to see which pair could get the highest score. Each rendition of the experiment involved two phases, loop creation by User 1 and loop replication by User 2. After a tempo is chosen (the tests generally stayed at bpm = 60), User 1 improvised a four measure gesture generated pattern in a 4/4 meter that continuously looped until User 2 felt they have internalized the pattern. User 2 then attempted to replicate the pattern and their rhythmic accuracy was scored on a scale of 0 - 100.

There were four levels of difficulty that originally corresponded to the number of gestures User 1 was permitted to use (Level I - one gesture, Level II - two gestures, etc.). This was later changed, after the initial testing (more on that in a later section), to four levels that corresponded to the number of measures in a pattern. Also, the absence or presence of visual feedback for User 2 vacillated to test the significance of the visual modality to the performance of the user. The participants within a pair interchanged the roles of User 1 and User 2 generating a total of 16 iterations per group.

Level I (1 measure)	Group (participants)	User 1	User 2	Visual	User 1	User 2	Visual	User 1	User 2	Visual	User 1	User 2	Visual
	1 (a, b)	a	b	yes	b	a	no	a	b	no	b	a	yes
Level II (2 measures)	Group (participants)	User 1	User 2	Visual	User 1	User 2	Visual	User 1	User 2	Visual	User 1	User 2	Visual
	1 (a, b)	a	b	no	b	a	yes	a	b	yes	b	a	no
Level III (3 measures)	Group (participants)	User 1	User 2	Visual	User 1	User 2	Visual	User 1	User 2	Visual	User 1	User 2	Visual
	1 (a, b)	a	b	yes	b	a	no	a	b	no	b	a	yes
Level IV (4 measures)	Group (participants)	User 1	User 2	Visual	User 1	User 2	Visual	User 1	User 2	Visual	User 1	User 2	Visual
	1 (a, b)	a	b	yes	b	a	no	a	b	no	b	a	yes

Figure 13: Outline of testing procedure.

Figure 14: Students using the system at the Escola de Música de Castellbisbal

4.2 Scoring

When User 2 replicates an existing gesture from the previously generated loop, his/her accuracy is scored as a weighted value between 0 and 1 depending on how late or early they played (a value of 1 being rhythmically perfect). They were penalized half of a point for playing when there was no gesture recorded in the loop.

To achieve this, User 1 generates a pattern and the performance of each gesture is stored into an array in Pure Data whose value at a given index is 1 for each recorded gesture and 0 for a rest. The performance is quantized to the eighth-note to prevent the generation of overcomplicated loops that would be difficult to replicate and score.

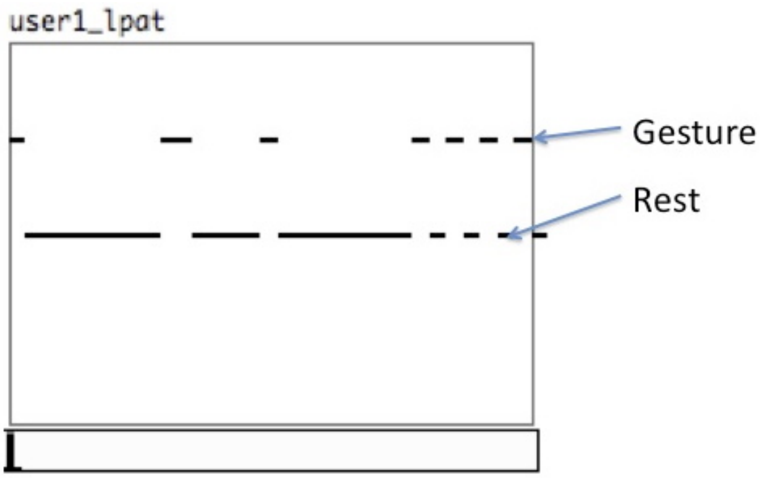
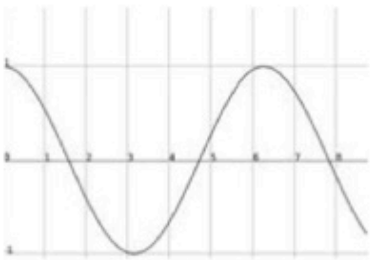


Figure 15: Pure Data array showing recorded performance for a gesture.

Each value then generates a corresponding waveform at a specific phase with a period value equivalent to one quarter note.

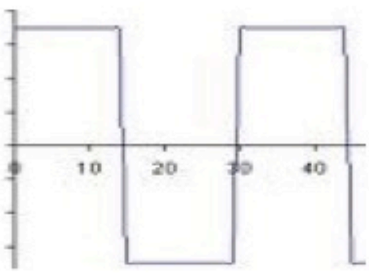
- Value of 1 generates a cosine function



$$A = 1, \theta = \frac{3\pi}{2}, T = \frac{60}{bpm}$$

Figure 16: The generated sine wave with its amplitude, phase, and period.

- Value of 0 generates square wave



$$A = -0.5, \theta = \frac{\pi}{2}, T = \frac{60}{bpm}$$

Figure 17: The generated square wave with its amplitude, phase, and period.

The recording of User 2's performance is un-quantized and the index value is computed using the amplitude of one of the two waveforms generated by the recorded performance of User 1, at the moment in time that a gesture is recognized from User 2. This method takes advantage of the higher rate and resolution of audio (44,100 Hz) in comparison to the control rate (1,000 Hz).

It was then necessary to create perceptual time shift for User 2 by delaying the feedback by a sixteenth-note, or $\mathrm{T} * 4$, so the peak, or trough, of the waveform lined up with the apex of the perceived beats, thus creating a dynamic weighted evaluation of the earliness, or lateness, of User 2's performance against the amplitude of the waveforms generated by User 1. The scoring is between 0 and 1 if User 2 plays on a gesture and they are given a -0.5 penalty for playing when there was no recorded gesture.

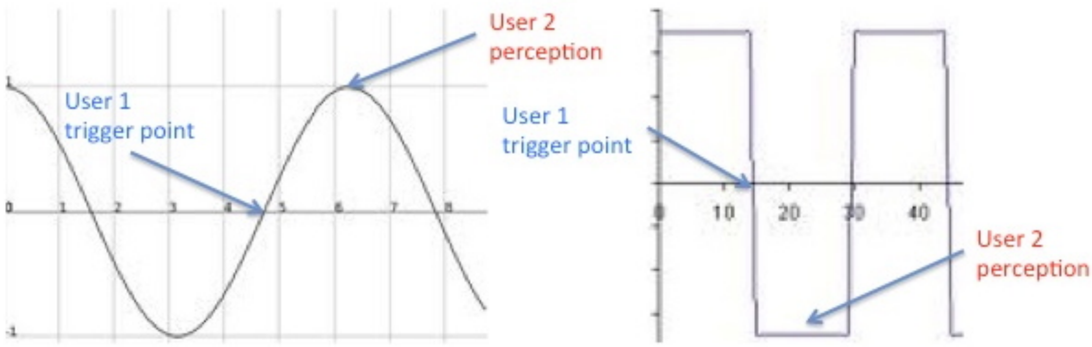


Figure 18: Perceptual time shift for each wave.

4.3 Evaluation

The system was tested on two occasions. During the first test with the students at the Escola Municipal de Música de Castellbisbal, user calibration issues prevented a proper quantitative evaluation. This was the first time the system was tested on multiple children at one time and there was difficulty tracking the skeletons of certain users. Investigation into the problem led to a rewrite of the source code which allowed for automatic user identification based on Center of Mass detection, the thinking being that if the detection algorithm did not require the user to adopt the calibration "psi"-pose a more fluid testing dynamic could be maintained [50].

The test did prove to be invaluable in the refinement of the testing method and system design. Working with the students' music professors, the number of gestures was brought down to two and the number of measures was made variable as to allow for the recording of patterns as short as one measure, or as long as four. These decisions were made because the task of replicating the loop proved too difficult for many in test group, possibly due to the fact that there was no real rehearsal time for User 2. The choice of the two gestures (left and right head pat) was provoked by the majority of the students in test group stating that performing that particular gesture was the most fun.

The second test was performed at a private home with four boys between the ages of seven and eight, with varying levels of music study. The system performed much better than the first test, and quantitative data was produced, but it was later discounted as continuing calibration issues caused the flow the game to be very slow and prevented the dynamic interaction necessary to maintain an active flow akin to a video game. It was decided that future developments in the system would forego Processing and investigate other software solutions, like the recent Kinect SDK distributed my Microsoft or FAAST. While Processing has the advantages of being multi-platform and a relative ease of use for non-programmers, its architecture causes it to be too computationally slow for a rhythmic performance application with the Kinect.

Observations from the second groups' interaction with the system provided useful vias for future development and improvements. The children were very excited to move in front the Kinect and trigger sounds but began to show a loss of interest being restricted to just two gestures. This desire to improvise suggested the need for a more open architecture with respect to gesture recognition that would allow for a broader library and/or more flexible interpretation of recognized gestures. The tests also proved to be too short when performed over one measure. After User 1 finished recording the initial pattern the children displayed an impulse to continue triggering sounds with their body and later expressed a desire to be able to layer sounds like a traditional step pattern sequencer. This input of design change led to the possibility of altering the focus of what could be measured to test viability, i.e. instead of measuring rhythmic accuracy we could measure rhythmic complexity to gauge if the system could aid in the development of rhythmic creativity [51].

5. Conclusions & Future Work

5.1 Conclusions

In this thesis an interactive gesture-based loop generation system has been presented as an aid to teach the concept of rhythm in the music classroom. It is the first step of a larger idea to provide tools for music pedagogy using inexpensive motion sensing technology in combination with open-source software. While the system itself did not perform as well as expected - mostly due to choice of software and the lack of calibration testing with children - the tests showed promising signs of viability. The children in both test groups were very motivated to use the system and excited to see the results of their performances. The professors from the first test group were also excited to see their students so motivated and saw potential in regularly using a system that engages the children in games which make use of concepts from the Orff method, like imitation and improvisation, while giving the teachers a tool for quantitatively tracking students' progress.

Combined with the research presented regarding the need for paradigm change in education, there is strong evidence supporting the development of teaching tools that are current and valid through the implementation of the technology utilized by our children every day. It is also clear that rhythmic embodiment and the use of body schema metaphors are essential elements in the effective teaching of abstract musical concepts, and any future developments of teaching systems would only benefit from the incorporation of these concepts in their design.

5.2 Future Work

Development and testing of future renditions of the system should be conducted over the course of a school year to gather data that can more concretely point to viability and pitfalls. The first step, with respect to design implementation, would be the investigation into improved calibration methods that could a give more seamless interaction experience to the user. Recently distributed open-source software and APIs developed for using the Kinect in subjects other than music could help in designing a more robust system, as well as allow for more flexibility with respect to recognized gestures. Also, advances in affordable depth camera technology now offer choices other than the Kinect, i.e. the Leap Motion 3D control system that promises 0.01ms of latency.

Overall, the research presented is encouraging and demonstrates the viability of conceptually similar systems that could be expanded to include the teaching of other musical concepts like volume, tempo, pitch, and timbre.

Bibliography

[1] K. Robinson, “Changing Paradigm,” *Royal Society for the Encouragement of Arts*, 2010. [Online]. Available: http://www.thersa.org/events/video/arch-robinson, (2010).ive/sir-ken.

[2] J. A. Aunión, “Recorte sobre recorte en educación,” *El País*, Madrid, 2012.

[3] J. A. Aunión, “Educación cambiará la ley para poder reducir profesores,” *El País*, Madrid, 2012.

[4] P. Díaz, “Educación propone subir un 20% el número máximo de alumnos por clase,” *El Mundo*, Madrid, 2012.

[5] B. Zamborlin, A. Sypniewski, N. Schnell, and N. Rasamimanana, “Continuous Realtime Gesture Following,”*Gesture in Embodied Communication and Human-Computer Interaction,* vol. 5934, pp. 73–84, Springer-Verlag, Hamburg, 2010.

[6] “Kinect Education.” [Online]. Available:http://www.kinecteducation.com/.

[7] Center For Digital Education, “Learning Through Motion: Increasing Student Engagement and Achievement with Immersive Technology,” Microsoft Corporation, 2012.

[8] F. Lehrdal and R. Jackendoff, *A Generative Theory of Tonal Music*. Boston: The Massachusetts Instute of Technology, 1983, pp. 12–36.

[9] M. Hefer, Z. Weitraub, and V. Cohen, “Music cognition: When does it start?,” *Early Childhood Music Education Commission 13th International Seminar of International Society for Music Education*, 2008, no. 8, pp. 35–39.

[10] J. Retra, “Music is Movement A Study into Aspects of Movement Representation of Musical Activities among Preschool Children,” Ph.D. dissertation, Dept. Education, University of Exeter, Exeter, UK, 2010.

[11] L. Habegger, “Number concept and rhythmic response in early childhood,” *Music Education Research*, vol. 12, no. 3, pp. 269–280, Sep. 2010.

[12] E. Church, *Music and Movement: A Practical Guide for Teaching Young Children*. New York, New York, USA: Scholastic, 1992.

[13] E. Orman, “Comparison of the National Standards for Music Education and Elementary Music Specialists’ Use of Class Time,” *Journal of Research in Music Education*, vol. 50, pp. 155–164, 2002.

[14] “National Standards for Music Education.” [Online]. Available: http://www.musiced.nafme.org/resources/national-standards-for-music-education/.

[15] J. Hurtienne and J. H. Israel, “Image schemas and their metaphorical extensions: intuitive patterns for tangible interaction,” in *Proceedings of the 1st international conference on Tangible and embedded interaction*, 2007, pp. 127–134.

[16] V. H. Mead, “More Than Mere Movement,” *Music Educators Journal*, vol. 4, no. 82, pp. 38 – 41, 1996.

[17] E. Jaques-Dalcroze, *Rhythm, Music and Education*. New York & London: G. P. Putnam’s Sons, 1921.

[18] J. Phillips-Silver and L. J. Trainor, “Feeling the beat: movement influences infant rhythm perception.,” *Science (New York, N.Y.)*, vol. 308, no. 5727, p. 1430, Jul. 2005.

[19] M. Leman, *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press, 2007.

[20] D. Goodkin, *Play, Sing and Dance An Introduction to Orff Schulwerk*. Mainz, Germany: Schott, 2004.

[21] P. Dourish, *Where the action is: the foundations of embodied interaction*. Cambridge, MA: MIT Press, 2001.

[22] J. Phillips-Silver and L. J. Trainor, “Hearing what the body feels: auditory encoding of rhythmic movement.,” *Cognition*, vol. 105, no. 3, pp. 533–46, Dec. 2007.

[23] A. N. Antle, M. Droumeva, and G. Corness, “Playing with the sound maker: do embodied metaphors help children learn?,” in *Proceedings of the 7th international conference on Interaction design and children*, 2008, pp. 178–185.

[24] S. Bakker and E. V. D. Hoven, “MoSo tangibles: evaluating embodied learning,” in *TEI’11*, 2011, pp. 85–92.

[25] S. Holland, A. J. Bouwer, M. Dalgelish, and T. M. Hurtig, “Feeling the beat where it counts,” in *Proceedings of the fourth international conference on*

Tangible, embedded, and embodied interaction - TEI ’10, 2010, pp. 25–27.

[26] W. Ahmad and A. Rambli, “Multi-sensory Modalities for Music Learning,” *Information Sciences*, pp. 1–5, 2008.

[27] S. Ferguson, “Learning musical instrument skills through interactive sonification,” in *Proceedings of the 2006 conference on New interfaces for Musical Expression*, 2006, pp. 384–389.

[28] M. Halpern, J. Tholander, M. Evjen, S. Davis, A. Ehlrich, K. Schustak, E. P. S. Baumer, G. Gay, and C. Ave, “MoBoogie : Creative Expression Through Whole Body Musical Interaction,” in *CHI 2011*, 2011, pp. 557–560.

[29] Y. Zhou, G. Percival, X. Wang, Y. Wang, and S. Zhao, “MOGCLASS : Evaluation of a Collaborative System of Mobile Devices for Classroom Music Education of Young Children,” in *International Conference on Human Factors in Computing Systems*, 2011, pp. 523–532.

[30] K. Takano and S. Sasaki, “An Interactive Music Learning System in Ensemble Performance Class,” in *2011 International Conference on Broadband and Wireless Computing, Communication and Applications*, 2011, pp. 65–74.

[31] C. Zhang, L. Shen, D. Wang, F. Tian, and H. Wang, “CoolMag : A Tangible Interaction Tool to Customize Instruments for Children in Music Education,” in *UbiComp’11*, 2011, pp. 581–582.

[32] S. Holland, P. Marshall, J. Bird, S. Dalton, R. Morris, N. Pantidi, Y. Rogers, A. Clark, and M. Keynes, “Running up Blueberry Hill : Prototyping Whole Body Interaction in Harmony Space,” in *Third International Conference on Tangible and Embedded Interaction*, 2009, pp. 93–98.

[33] T. D. Stanley, B. Y. University-hawaii, and D. Calvo, “Rhythm Learning with Electronic Simulation,” in *SIGITE’09*, 2009, pp. 24–28.

[34] K. Ng, “Technology-enhanced learning for music with i-maestro framework and tools,” in *EVA London 2008: Electronic Visualisation & the Arts. Proceedings*, 2008, pp. 177–187.

[35] Cycling ’74, “Max/MSP,” [Online]. Available:http://cycling74.com/products/max/.

[36] F. Bevilacqua, F. Guédy, N. Schnell, E. Fléty, and N. Leroy, “Wireless sensor interface and gesture-follower for music pedagogy,” *Proceedings*

of the 7th international conference on New interfaces for musical expression, 2007, pp. 124–129.

[37] A. N. Antle, G. Corness, and M. Droumeva, “Human-computer-intuition ? Exploring the cognitive basis for intuition in embodied interaction,” *Int. J. Arts and Technology*, vol. 2, no. 3, pp. 235–254, 2009.

[38] M. Banzi and D. Cuartielles, “Arduino.” [Online]. Available: http://www.arduino.cc/.

[39] “Microsoft Kinect.” [Online]. Available:http://www.microsoft.com/en-us/kinectforwindows/ .

[40] play.com, “Kinect Product Page.” [Online]. Available: http://www.play.com/Games/Xbox360/4-/10296372/Project-Natal/Product.html.

[41] H. J. Hsu, “The Potential of Kinect as Interactive Educational Technology,” *IPEDR*, vol. 13, pp. 334–338, 2011.

[42] “Simple Openni.” [Online]. Available:http://code.google.com/p/simple-openni/.

[43] B. Fry and C. Reas, “Processing.” [Online]. Available:http://processing.org/.

[44] “Prime Sense.” [Online]. Available:http://www.primesense.com/.

[45] G. Borenstein, *Making Things See*. Gravenstein Highway North, Sebastopol, CA: O’Reilly Media, Inc., 2012.

[46] B. Buxton, “NUI with Bill Buxton," *CES,* 2010, [Online]. Available: http://channel9.msdn.com/Blogs/LarryLarsen/CES-2010-NUI-with-Bill-Buxton

[47] N. P. Lago and F. Kon, “The Quest for Low Latency ,” *Proc. of the 2004 Int. Computer Music Conf,* 2004, pp. 33-36.

[48] J. L. Chen, V. B. Penhune, and R. J. Zatorre, “Moving on Time : Brain Network for Auditory – Motor Synchronization is Modulated by Rhythm Complexity and Musical Training,” pp. 226–239, 2008.

[49] “TouchOSC.” [Online]. Available: http://hexler.net/software/touchosc.

[50] A. A. Manasrah, “Human Motion Tracking for Assisting Balance Training and Control of a Humanoid Robot,” M.S. thesis, Dept. of Mechanical Engineering, University of South Florida, Tampa, FL, 2012.

[51] J. Pressing, “Cognitive complexity and the structure of musical patterns,”*Noetica,* vol. 3, no. 8, pp. 1–8, 1998.