A Preliminary Evaluation of the Leap Motion Sensor as Controller of New Digital Musical Instruments

Eduardo S. Silva[1], Jader Anderson O. de Abreu[1], Janiel Henrique P. de Almeida[1], Veronica Teichrieb[1], Geber L. Ramalho[1]

[1]Centro de Informática – Universidade Federal de Pernambuco (UFPE) Caixa Postal 50.740-560- Recife – PE – Brasil.

{ess,jaoa,jhpa,vt,glr}@cin.ufpe.br

Abstract. *The introduction of new gesture interfaces has been expanding the possibilities of creating new Digital Musical Instruments (DMIs). Leap Motion Controller was recently launched promising fine-grained hand sensor capabilities. This paper proposes a preliminary study and evaluation of this new sensor for building new DMIs. Here, we list a series of gestures, recognized by the device, which could be theoretically used for playing a large number of musical instruments. Then, we present an analysis of precision and latency of these gestures as well as a first case study integrating Leap Motion with a virtual music keyboard.*

1. Introduction

Human-Computer Interaction (HCI) is an important field of study that has gained increasing attention with the emergence of new interaction devices (such as de Nintendo Wii Remote[1] and the Microsoft Kinect[2]). The use of gestural interfaces are arousing interest, since interaction is essential in several domains of application such as art, medical assistance, simulation, etc. [Freitas et. al. 2012; Singer, Larke and Bianciardi 2003; Wong et. al. 2008].

In this context, the computer music field has been witnessing a growing number of new Digital Musical Instruments (DMIs), that have benefit from new interface technologies [Miller and Hammond, 2010; Montag, Sullivan, Dickey and Leider, 2011; Jordà, Kaltenbrunner, Geiger and Bencina 2005].

Several of these DMIs can be found all over internet sites like Youtube[3] and Vimeo[3]. DMIs such as the Kin Hackt[4] developed by Adriano Clemente and the work of Chris Vik in collaboration with the dancer Paul Walker[5], integrate the Kinect sensor

[1] Wii Remote is a motion controller from the Nintendo© Wii Console developed by Nintendo© Company, Limited. Nintendo© Wii website: http://www.nintendo.com/wii.

[2] Microsoft© Kinect is a depth sensor developed by Microsoft© Corporation initially for the Microsoft© Xbox 360 videogame console to be a motion controller, then updated to function with Microsoft© Windows computers. Microsoft© Kinect websites: http://www.xbox.com/en-US/kinect and http://www.microsoft.com/en-us/kinectforwindows/.

[3] Youtube and Vimeo are video-sharing websites that allow users to upload, view and share videos.

[4] Video available at http://www.youtube.com/watch?v=YW86yyz0gj0.

[5] Video available at http://www.youtube.com/watch?v=qXnLxi2nzrY.

capabilities with the software Live[6] by Ableton to enable musical composition. Projects like the V motion project[7], built a DMI that uses two Kinect sensors to enhance precision and reduce latency at the same time as giving visual feedback for artistic purposes.

New instruments using other videogame motion controllers such as the Wii remote (Silva, 2012; Miller and Hammond, 2010), projects using multi-touch tablets [Calegario, 2013], multi-touch tables [Jordà, Kaltenbrunner, Geiger and Bencina, 2005] or building their own innovative DMIs such as Mike Waisvisz's "Hands"[8], developed in the 1980's, have also been explored.

Moreover, several researchers have addressed the evaluation of new technologies and DMIs. Studies evaluated precision in musical properties such as tempo, latency and precision with instruments built with the Wii remote controller alone [Kiefer, Collins, and Fitzpatrick, 2008], with the Wii sensor bar [Peng and Gerhard, 2009] and with several multi-touch platforms available at the time [Montag, Sullivan, Dickey and Leider, 2011]. Others approached piano movement analysis with inertial sensors attached to the user [Hadjakos and Mühlhäuser, 2010].

Since the area is relatively new, there is a lack of consolidated method to evaluate DMIs. Therefore, researches have used different methods to approach this problem, varying from using methodologies based on HCI concepts and theories [Wanderley, and Orio 2002] to a more direct analysis focusing on comparison of acoustic and digital instruments performances [Collicutt, Casciato and Wanderley, 2009].

In 2012, the Leap Motion controller was introduced [Leap Motion inc., 2012]. This gesture controller provides an approximately 150º field of view and uses a depth sensor to track hand features up to 1/100th of a millimeter. This fine-grained control may represent an opportunity to create a new generation of DMI. However, to confirm Leap Motion potential, an evaluation should be performed concerning latency and precision, some of the common bottlenecks in the use of gestural interfaces in DMIs [Silva 2012; Costa Júnior et. al. 2011].

In this paper, we perform a preliminary study and evaluation of the Leap Motion sensor as a tool for building DMIs. We start by listing the conventional music gestures that can be recognized by the device. Then, we analyze precision and latency of these gestures. As part of the evaluation method, we also propose a first case study integrating Leap Motion with a virtual music keyboard, which was called "Crystal Piano".

2. Leap Motion Controller

A DMI is an interactive artifact used for musical purposes that separates the input (gesture control) from the output (sound synthesis) using mapping strategies to associate them [Malloch, Birnbaum, Sinyor, and Wanderley, 2006; Calegario, 2013].

In the last 10 years, various projects illustrate the impact of new interfaces in building DMIs [Silva 2012; Jordá 2005; Wong et. al. 2008]. Leap Motion is one of these recent technology advances that may have an impact on the creation of DMIs.

[6] Live is a digital audio workstation created by Ableton. https://www.ableton.com/en/live/.

[7] The V motion project video is available at http://vimeo.com/45417241.

[8] "Hands" is a Digital Music Instrument developed at STEIM (steim.org). Description available at http://www.crackle.org/TheHands.htm.

The Leap Motion works with two infrared (IR) cameras and three IR LEDs as a depth sensor in a limited field of view (FOV) (Figure 1) of 8 cubic feet (approximately 61 cubic centimeters). Using the stereoscopy from both cameras, the device can minimize errors from tools, fingers and hand features and is built on a unique mathematical model to maximize speed and precision. As the device detects these features, it provides updates in frames of data. Each frame has a list of tracking data such as hands, fingers, tools, recognized gestures and factors describing the overall motion of the scene.

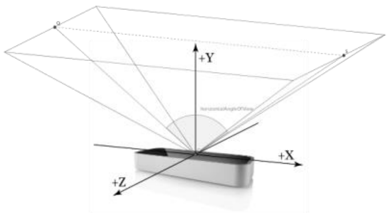

Figure 1. Leap Motion's Field Of View.

This allows Leap Motion to recognize hand features such as hand palm orientation, fingers’ length, width and orientation, hand opening and other non-hand features including tools and computer screen location. Furthermore, it incorporates in its Software Development Kit (SDK) the recognition of a few gestures including “circle gesture”, swipe and tap that are shown in its interface (Figure 2). With respect to controllers like Nintendo Wii Remote and Microsoft Kinect, more focused on body and body members, Leap Motion provides a fine-grained hand control, which is clearly promising for building new DMIs.

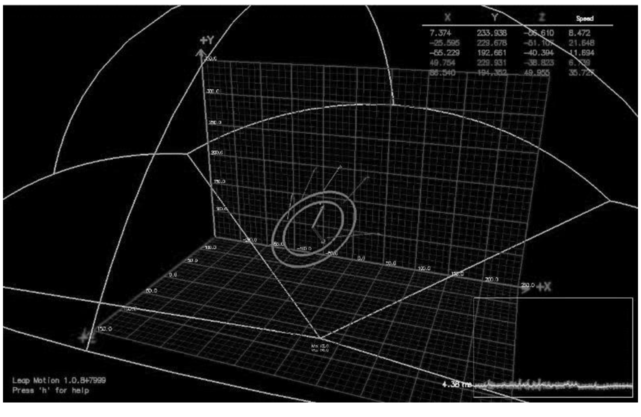

Figure 2. Leap Motion SDK’s FOV and tracking features.

Although it has a limited space to work with, the hand tracking is very precise. The Leap Motion SDK comes with three different presets of tracking: Robust mode, Low resource mode and Auto orientation device mode. Furthermore, it has also three different performance sets: High precision, High Speed and Balanced tracking.

In the High Precision mode, the dataflow provides circa 50 data frames per second (data fps), representing about 20 milliseconds (ms) of delay in the computer music field. With the Balanced tracking its data fps increases by a factor of two, reducing the delay to 10 ms, combined with a still good precision. By choosing the High Speed mode, it loses a perceptible amount of precision with tracking but reduces the delay to 5 ms (approximately 200 fps).

3. Methodology

Considering that the musical field is highly demanding of time precision, low latency is a parameter to be considered when building a new instrument. In digital music instruments, this problem is even more critical, since the gesture control has to be processed by a computer that will perform a series of calculations, then will synthetize a sound, trigger an event or continually manipulate sound properties or processing.

Taking into account that triggering an event requires a lower latency than a continuous sound manipulation [Calegario, 2013], we decided to evaluate the more critical one. Therefore, measuring the latency in which the sensor can capture gestures and trigger an event is an important parameter to evaluate a new technology for building DMIs.

With that in mind, we decided to perform a series of gestures and measure the latency thereof. A simple software was developed so when a finger surpasses a certain threshold a sound was played. A glass table placed at the threshold's position interrupted the movements. To capture the sound of the finger hitting the table, a microphone was placed on it. The delay measure is the time difference between the sound of the finger hitting the table and the sound played by the software.

Furthermore, a simple DMI was built allowing us to evaluate latency and precision of multiple events being triggered sequentially or simultaneously.

After an analysis with images, videos, articles, and the device's documentation, the study of possible applications was made. After choosing to simulate a simple piano, different interfaces and tools were analyzed to identify which of those available in the literature would be appropriate to integrate with the Leap Motion Controller.

Among several interfaces, the Synthesia©[9] game [Synthesia© LLC., 2012], was identified as the most promising one. Using gamification aspects for teaching users to play piano, it functions, naturally, with any MIDI keyboard or device, however it also can function with the conventional QWERTY computer keyboard. The software is able to stop, resume, slow down and speed up the song's rhythm and can play any MIDI song.

However, this former open-source interface has turned private and its code could not be used.

Another attempt to use this interface was trying to simulate the QWERTY keyboard with the Leap Motion. But it brought up two problems. The first one was that, the time of a musical note had to vary with the pressed key, and the synchrony was not ideal. The other problem was due to an Operation System (OS) aspect. When dealing with multi-processing, the application that is being used acquires a focus and has priority of execution, therefore the input from the QWERTY keyboard and the device, goes straight to the Synthesia, without being processed by the application that emulates the Leap Motion, i.e., the system's architecture was not helpful.

[9] The Synthesia© software is a music educational tool developed by Synthesia©, LLC. Description at: https://www.synthesiagame.com.

We, then, built our own interface that simulates one octave of a piano and evaluated the DMI.

4. Precision and Recognition Studies

4.1. Gesture Mapping

After understanding the articulations and movements of the upper body members [Gray 1918], it is possible to understand which of those are responsible for accomplishing the musical gestures. It is necessary to emphasize that most of musical gestures are made by using the upper body extremities (shoulders, arms, elbows, forearms, wrists and hands). Previous studies classified those movements by type of articulation and highlighted its Degrees Of Freedom (DOFs) [Silva 2012].

Particularly, the piano is an instrument that requires short and precise hand gestures to be played. With that in mind, the classification of the interphalangeal articulations was taken as an additional study of the digits:

- Proximal interphalangeal articulation: Articulation between the first and second phalanges (1 DOF).
- Distal interphalangeal articulation: Articulation between the second and the third phalanges (1 DOF).

The possible movements for these articulations are only flexion and extension (therefore 1 DOF) and are more extensive in the proximal articulations. It is important to highlight that the thumb has only one interphalangeal articulation, while the remainder of the fingers (on the hand) have two.

Therefore, combining the movements listed in previous studies and the detail of digits articulations, it is able to reproduce the gestures for playing the piano. Follows a list thereof, according to Table 1:

Table 1. Upper body movements to play a piano.

Movement	Recognition
Shoulder Movement	X
Arm Movement	X
Elbow Movement	X
Forearm Movement	X
Wrist Movement	X
Metacarpo-Interphalangeal Movement	X
First Interphalangeal Movement	X
Second Interphalangeal Movement	X

4.2. Experiments and Results

After the study of the possible gestures, it was time to analyze which of these the tracking sensor in study can recognize. The device was tested using the version v0.8.1.6221 of the SDK. The test included a series of hand manipulations, articulations movements and gestures similar to those required to play the selected instrument.

The device can capture hand features as palm position, orientation and aperture, fingers length, width and orientation. Although it cannot capture the position of the junction points and segments from the whole arm, it can accurately detect the motion passed into the hands, according to Table 2.

Table 2. Piano movements captured by the Leap Motion Controller.

Movement	Recognition
Shoulder Movement	X
Arm Movement	X
Elbow Movement	X
Forearm Movement	X
Wrist Movement	X
Metacarpo-Interphalangeal Movement	X
First Interphalangeal Movement	X
Second Interphalangeal Movement	-

After the study of which gestures the sensor can capture, the simple process to measure the latency thereof was made.

When calculated, the delay for the recognition was approximately of 71 milliseconds (ms) in the High Speed mode. According to [Jordà, 2005], an acceptable latency regarding triggered events has to be lower than 20 ms, so the latency found is unacceptable for triggering events.

5. Latency Study

5.1. Case Study

The developed interface (Figure 3) simulates one octave from a piano and allows the user to play it using the Leap Motion controller. This interface shows the user where his fingers are in space.

Figure 3. Crystal piano interface.

5.2. Usability Aspects

In matter to provide tactile feedback to the user, it was created a set-up with a glass “table” above the device (Figure 4). To simulate the keys from the keyboard, the glass was marked in the same way a piano should look.

Figure 4. Piano DMI setup with Leap Motion controller under the glass table.

To provide better usability to the system, hand and body postures from online piano educational classes were analyzed. Usability is defined as the ability of an interactive system offers its user in a given context of operation, to carry out tasks in an efficient, effective and enjoyable [ISO 9241, 1993].

Thus, we analyze how the position to play a piano should be. This will help you understand which aspects should be kept to the user to make a good use of Crystal Piano.

The posture is influenced by factors such as:

- The biotype of each user and his body size;
- The position to play standing or sitting, since this may influence respiration;
- The chosen way of supporting the instrument, since this will determine the position of arms and lips;
- The curvature of the nozzle, since it affects the angle at which the blade will be in relation to the lips.

These factors need to be observed to obtain an efficient and comfortable position to play the piano.

It is noteworthy that, in addition to good posture, the position of the hands and their indications are references in the act of playing. To do so, the fingers are numbered for easy positioning in writing scores, where the numbering is the same for the right and left hands.

5.3. Implementation Details

The software used to perform the case study was developed in Java to simulate a music keyboard. It uses gesture input associated with a MIDI protocol to simulate the sound.

The software was created using the following libraries:

- LeapMotion: Official library of the Leap Motion sensor that allows the program to capture the sensor's data information and requires a pre-configured structure to identify the gestures and tools recognized by the device.
- LWJGL: Java library that gives access to high performance cross-platform libraries, such as OpenGL, OpenCL and OpenAL.
- SLICK: Java library compatible with LWJGL to facilitate the development of 2D games.

5.4. Experiments and Results

As a first case study, the user experienced playing different songs. Some aspects were analyzed:

1. Latency of the output.
2. False positive input (the note is played, but the user did not played it).
3. False negative input (the note is not played, but the user did played it).
4. Presence of tactile feedback set-up.
5. Software Interface.
6. Posture.

Follows, the user's perspective (Figure 5):

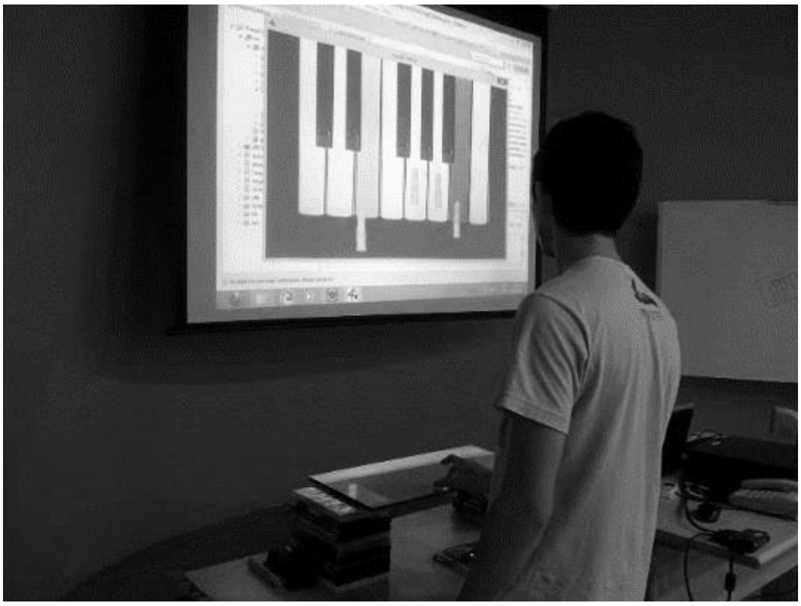

Figure 5. User's perspective.

Regarding physical limitations of the current set-up, the interface only shows one octave (from Do C4 to Do C5).

By using a projector system, the user was able to follow the hand tracking. Virtual fingers provided visual feedback to the user.

The tactile feedback was provided by mounting a glass table above the device as shown in (Figure 6). Markers were used to provide visual feedback on the glass. The table's height was adjusted by an improvised stacking of boxes.

Figure 6. User's perspective of the setup.

The user played three different songs and the following aspects were noted:

- All fingers appear (on the interface) as the pointing finger, losing fidelity;
- The interface shows were the fingers are;
- It appears to be intuitive, in the way that when the user places the hand above the glass, the interface shows the fingers' positions. However, an user that does not have the musical skill to play the piano, cannot perform a complex song (as with a real piano);
- Only displays one octave;
- The latency measure with the render interface was made using the same method as with the gestures recognition and presented a large latency;
- When a song was played fast, the user had problems related to false positive input, false negative input and latency;
- The presence of the glass table was great when it comes to having something to touch and to be able to know where is the sound trigger;
- The user could move his hand freely, but when occlusion and tracking failure factors were considered, the gesture freedom had some constraints;
- When two fingers were too close, the tracking was lost and then recovered, causing false positive input.

The major problem was related to the tracking. When two fingers were too close to each other, the Leap Motion cannot resolve them, losing tracking. In piano gestures, when the user is playing a song, his fingers can be close at some point. If the finger is pressing one note, the system, by losing tracking and recovering it, plays another note or a false positive input. However, if the system does not capture the pressed finger, no note is played and the user experiences a false negative input.

Occlusion plays a great part in the errors as well. The player has to adjust the hand position in a way to the system to identify all the fingers.

When playing a song in a fast way, the user experienced some tilt from the interface. Combining the fast tracking with occlusion and tracking loss, the system, sometimes, tilts causing frustration in the user. However, when the user plays a song in a slow way, the system flows well.

The presence of a tactile feedback was very enjoyable to the user, which experienced playing songs with and without the glass table. Playing without the table can cause loss of the key's position, forcing the user to suppose it. With the glass, the trigger position is physically displayed and the user gains a more precise control. In addition, the markings on the table provides visual feedback of which piano key is pressed. The down side is that there is no actual key (it is a touch key) nor a force feedback.

Another good point of the tactile feedback set-up is that the user "rests" his hand on the table, since there is a contact to overcome gravity force applied in the hands (normal force). When playing without the set-up, the user experienced some discomfort from hanging his arms and hands in the air for some time.

7. Conclusions

With this study, the analysis of the device shows that it is a powerful tool for recognizing hand gestures with great precision. The SDK showed a dataflow that produces a delay from 5 ms up to 20 ms, with a middle level of 10 ms with great precision. The latency of recognition measured with the prototype developed in Java code presented a still unacceptable delay to build DMIs with precise hand tracking. However, a further study with this device has to be made to verify the performance with other code languages (such as C++) and algorithms.

The case study system proved to be a good simulation only when played in a slow rhythm. In normal-fast paced songs, the system presented an unacceptable latency. The occlusion and loss of tracking problems with the current SDK are relevant, but the Leap Motion proved to be a powerful tool to simulate musical instruments, by having a great gesture repertoire. Meanwhile, the presence of a tactile feedback was of very importance to the user's experience.

The piano, by requesting a particular hand precision, may not be the ideal instrument to be simulated with only one depth sensor.

Concerning future works, the analysis of other music gestures using the device itself or its tracking principle has to be made. Furthermore, the analysis of other music instruments and their gestures is of utmost importance to investigate more the device's capabilities.

It is also important to test other code languages and algorithms to verify the device's performance. Furthermore, it is extremely important to study the integration of this device with other technologies, a work that is already in progress.

A thorough analysis of the prototype's software and improvement of thereof has to be made in order to correct the software's problems and to simulate more than one octave. Another possibility is the development of a driver that, once installed, captures the sensor's data and turns them into a MIDI input signal or that simulates a QWERTY keyboard smoothly. That way, it will be possible to integrate the Leap Motion controller with other tools like the Synthesia© Game or even other applications like a text editor.

8. References

Calegario, F. C. A., (2013).Sketchument: ambiente de experimentação para criação de instrumentos musicais digitais. Dissertação de mestrado. Universidade Federal de Pernambuco, Recife.

Collicutt, M., Casciato C. and Wanderley, M. M. (2009). From Real to Virtual: A Comparison of Input Devices for Percussion Tasks**,** Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 1–6.

Costa Júnior, J. B., Calegario, F. C. A., Magalhães, F., Cabral, G., Teichrieb, V. and Ramalho, G. L. (2011). Towards an evaluation methodology for digital music instruments considering performer s view: a case study.In: Simpósio Brasileiro de Computação Musical, 2011, Vitória - ES. Anais do 13º Simpósio Brasileiro de Computação Musical.

Freitas, D. Q., Gama, A. E. F., Figueiredo, L., Chaves, T. M., Oliveira, D. M., Teichrieb, V. and Araujo, C. (2012) Development and Evaluation of a Kinect Based Motor Rehabilitation Game. Simpósio Brasileiro de Jogos e Entretenimento Digital, 2012, Brasilia. SBC – Proceedings of SBGames 2012.

Gray, H. (1918). Anatomy of the Human Body. Disponível em <http://www.bartleby.com/107/>. Acessado dia 30 de agosto, 2013.

Hadjakos, A. and Mühlhäuser, M. (2010). Analysis of Piano Playing Movements Spanning Multiple Touches. Proceedings of the International Conference on New Interfaces for Musical Expression. Retrieved from http://www.nime.org/proceedings/2010/nime2010_335.pdf.

ISO 9241 Part 1(1993). Ergonomic requirements for office work with visual display terminals, Part 1 General Introduction; International Standard ISO 9241-1.

Jordà, S. (2005). Digital Lutherie. Universitat Pompeu Fabra. Disponível em <http://dialnet.unirioja.es/servlet/tesis?codigo=19509>. Acessado dia 30 de agosto, 2013.

Jordà, S., Kaltenbrunner, M., Geiger, G. and Bencina R. (2005). The reactable. Proceedings of the international computer music conference (ICMC 2005), Barcelona, Spain, pages 579-582. Citeseer.

Kiefer, C., Collins, N. and Fitzpatrick, G. (2008). HCI methodology for evaluating musical controllers: A case study. Proceedings of the International Conference on New Interfaces for Musical Expression. Retrieved from http://nime2008.casapaganini.org/documents/Proceedings/Papers/193.pdf.

Leap Motion Incorporation. Retrieved from: https://leapmotion.com/. 29 de October 2012.

Malloch, J., Birnbaum, D., Sinyor, E. and Wanderley, M. M. (2006). Towards a New Conceptual Framework for Digital Musical Instruments. Proceedings of the 9th International Conference on Digital Audio Effects (pp. 49–52). Retrieved from http://www.dafx.ca/proceedings/papers/p_049.pdf.

Miller, J. and Hammond, T. (2010). Wiiolin: a virtual instrument using the Wii remote. Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010), Sydney, Australia.

Montag, M., Sullivan, S., Dickey, S. and Leider, C. (2011). A Low-Cost, Low-Latency Multi-Touch Table with Haptic Feedback for Musical Applications. Proceedings of the International Conference on New Interfaces for Musical Expression, (June), 8–13.

Peng, L. and Gerhard, D. (2009). A Wii-based gestural interface for computer conducting systems. Proceedings of the International Conference on New Interfaces for Musical Expression. Retrieved from www.nime.org/proceedings/2009/nime2009_155.pdf.

Silva, J. V. S., (2012). Avaliando Interfaces Gestuais Para Prática de Instrumentos Virtuais de Percussão. Dissertação de Mestrado. Universidade Federal de Pernambuco, Recife.

Singer, E., Larke, K. and Bianciardi, D. (2003). LEMUR GuitarBot: MIDI Robotic String Instrument. Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03), Montreal, Canada.

Synthesia Game LLC. Retrieved from: https://www.synthesiagame.com. 29 October 2012.

Wanderley, M. M., and Orio, N. (2002). Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI. Computer Music Journal, *26*(3), 62–76.

Wong, E., Yuen, W. and Choy, C. (2008). Designing wii controller: a powerful musical instrument in an interactive music performance system. In Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia, pages 82-87. ACM.