MuDI - Multimedia Digital Instrument for Composing and Performing Digital Music for Films in Real-time

Pedro Patrício
Research Centre for Science and Technology of the Arts (CITAR)
Portuguese Catholic University - School of the Arts
Rua Diogo Botelho 1327, 4169-005 Porto-Portugal
ppatricio@porto.ucp.pt

ABSTRACT
This article proposes a wireless handheld multimedia digital instrument, which allows one to compose and perform digital music for films in real-time. Not only does it allow the performer and the audience to follow the film images in question, but also the relationship between the gestures performed and the sound generated. Furthermore, it allows one to have an effective control over the sound, and consequently achieve great musical expression. In addition, a method for calibrating the multimedia digital instrument, devised to overcome the lack of a reliable reference point of the accelerometer and a process to obtain a video score are presented. This instrument has been used in a number of concerts (Portugal and Brazil) so as to test its robustness.

Keywords
Digital musical instrument, mobile music performance, real-time musical composition, digital sound synthesis.

1. INTRODUCTION
In response to an invitation to play live music for a film we decided to design MuDI, a wireless multimedia digital instrument to compose and perform music in real-time.

The MuDI is a multimedia instrument because it simultaneously (1) diffuses sound, image and video, (2) operates sound and video in real-time, (3) records the sound of the performance, (4) allows to follow not only the film's images, but also the relationship between the gestures performed and the sound generated, and (5) at the end of the performance allows to obtain an audiovisual file (composition) in addition to a video score.[1]

One of the goals of this work is to achieve great musical expression. Hence, we felt the need to obtain a stable and reliable musical system as well as an effective control over the sound. A response to these needs relies on the process of calibration that allows to overcome the lack of a reliable reference point of the accelerometer [1] (the main sensor of interaction used for playing the MuDI) in addition to a laptop computer application interface that allows to follow the relationship between the gestures performed and the sound generated.

Furthermore, the fact of having to perform live, in real-time and with quite precision the music previously composed, the need to have a musical guide or score was also felt. We will also present a method to automatically obtain a video score, which is useful to compose and perform music in real-time.

MuDI was used to compose "Sonification - films for music$1", a work premiered in September 2009, in the 12th Brazilian Symposium of Computer Music (SBCM).[2] Recently, we created a new composition titled "Sonification - films for music$2" (2012), in which we use images extracted from the film Metropolis (1927) by Fritz Lang.[3] The last presentation of MuDI took place in October 2011, in the 2nd International Forum of Musical Itineraries - Music and Gesture, Faculty of Social Sciences and Humanities, Universidade Nova de Lisboa, Lisbon, Portugal.[4]

Related work with MuDI, a more detailed description of its system, handling, calibration and composition processes will be discussed in the following sections of this article.

2. RELATED WORK
The conception of the MuDI is based on previous works, which explore the possibility of using handheld devices, such as personal digital assistant (PDA) and smartphones, as controllers and musical instruments.

Geiger [2], [3] explored the possibilities of using a single touch screen of a PDA as a musical controller. For this purpose he implemented a user friendly interface constituted by a Pure Data FM synthesiser, a virtual guitar and a virtual drum set. In this project the user operates the interface either with a stylus or directly with the fingers. Hence, gestures such as pressing, strumming and tapping are used to achieve musical expression. However, as the touch screen does not support multi-point detection, i.e. only one detection point at a time is allowed, it is difficult to attain musical expression.

New approaches, such as the attachment of external sensor units to the handheld devices allowed the users to extend the possibilities of interaction and consequently to create a new gestural repertoire.

Following this approach Tanaka [4] expanded the PDA with an external sensor unit constituted by a force sensing resistors to capture the grip pressure, an accelerometer to capture the position and the velocity of the motion of the user's gestures, as

[1] A promotional video about the MuDI could be seen at https://www.youtube.com/watch?v=leh-FMwIh5g

[2] View two video excerpts of the concert (excerpt1 and excerpt2) at https://sites.google.com/site/pp2007pt/uk/mudi

[3] This digital musical composition could be listened at https://www.youtube.com/watch?v=rg5Kw2LMsDQ

[4] Two video excerpts of the presentation could be seen at https://sites.google.com/site/pp2007pt/uk/mudi

well as a Wi-Fi network card in order to transmit the audio between the users. In this project, the musical expression is achieved in the following manner: The intensity of the force applied by the user when grabbing the PDA transforms the brightness of the sound, the tapping of the rhythm controls the re-sequencing of the song in the time-domain and when the users get geographically closer to each other, the individual parts of the song, previously selected by them become prominent in the final mixture which is being shared through the Wi-Fi network.

Other technological alternative means, such as the interactions based on integrated camera phones were tested in projects as CaMus [5]. In CaMus, the user moves a camera phone over a surface grid of visual markers. The data extracted from the position, distance and angles of rotation of the camera in relation to the markers are then converted into sounds through a music sequencing software. However, the continuous motion over the surface grid promotes the constant change of sound thus making it difficult to find the right pitch. The low capability of the camera to capture fast movements causes a perceptible delay between the gesture performed and the sound generated. Additionally, the limited extension of the surface grid and the extremely short recognition distance between the markers and the lens of the camera limits the use of free gestures for interacting [1], and subsequently make it difficult to achieve the desired musical expression.

The development of accelerometers and its integration in common handheld devices have provided the users the possibility to perform gestures in an open space, i.e. the users' gestures for interacting with the handheld devices are no more limited to the size of a unique surface such as in the case of single touch screen used in Geiger [2], [3] or in the grid surface used in CaMus [5]. Furthermore, the capability of the accelerometers to capture the movements in a fast way allows to attenuate the perception of delays between the gestures performed and the sound generated. The accelerometers can be used for very precise and fast motions [1].

From this type of built-in sensors, a new vocabulary of gestural interaction has emerged for achieving musical expression in a simpler and more precise way. For example, gestures used for playing traditional musical instruments (drums, rattles and violin) such as striking, shaking and sweeping were used in ShaMus for playing a mobile phone [6].

In the case of MoGMI [7], the user moves a mobile phone along the 3-axes of the accelerometer. The user's recorded movements are later translated into MIDI commands and subsequently played back using an on-board MIDI player. Each axis controls a different musical parameter: The x-axis manipulates the attack of the sound, the y-axis the pitch and the z-axis the intensity of the sound. However, the use of an offline MIDI recorder/player for translating recorded movements into MIDI messages does not allow to play music in real-time [7].

The integration and the combination of both technologies, the multi-touch screen (which allows multi-point detection) and the accelerometer in the same handheld device (e.g. the Apple's iPod touch) has expanded even more the possibilities of interaction. Subsequently, musical expression can be achieved in a more efficient way.

These new possibilities were exploited in projects such as MoPho [8], a mobile phone orchestra for composing and performing digital music. The detection of the position and the speed of the movement of the device through the accelerometer and the multi-point detection through the multi-touch screen are the principal means of interaction used in MoPho to achieve musical expression.

When interacting with handheld devices based exclusively on touch screens, generally both hands and the eyes are required to operate them: One hand to hold the device and the other to operate it. In addition, looking is indispensable not only to see the user's interface but also to control its handling.

On the contrary, operating an accelerometer requires only one hand (the hand that holds the device) and one has not to look to the device in order to control it. However, in relation to the multimedia instrument that is being proposed in this article, this latter aspect is essential, in a sense that by freeing the eyes at the device the user is available to follow not only the film images but also the relationship between the gestures performed and the sound generated through the laptop computer application interface that is projected during the performance (see Figure 2).

3. MuDI SYSTEM

The MuDI's system is constituted by a handheld device, a laptop computer, a controller application and a laptop computer application interface.

Seeing that contemporary mobile handheld devices are easy to operate and have more integrated sensors than laptop computers, the iPod touch was chosen as our exclusive expression medium and musical controller.

On the other hand, a laptop computer was selected to function as a sound generator, due to the fact that nowadays, they have a greater capacity for processing information, storage and memory than the more common PDAs or smartphones. In addition, they have a wide variety of versatile and powerful programs focused on the generation of digital sound synthesis (e.g. Max/Msp, Pure Data, SuperCollider, Csound, Chuck).

Furthermore, the MuDI requires the use of at least one laptop computer because it requires the use of a build-in Wi-Fi network technology as well as an audio and video output. The wireless connection between the handheld device and the laptop computer is established through of the OpenSound Control (OSC) Protocol [9].

There are several controller applications for the iPod touch, among which the most flexible is the OSCRemote.[5] It is an intuitive controller that allows easy creation and editing of user control interfaces within the application using graphical user interface widgets, such us buttons, switches, sliders and accelerometers (see Figure 1 below).

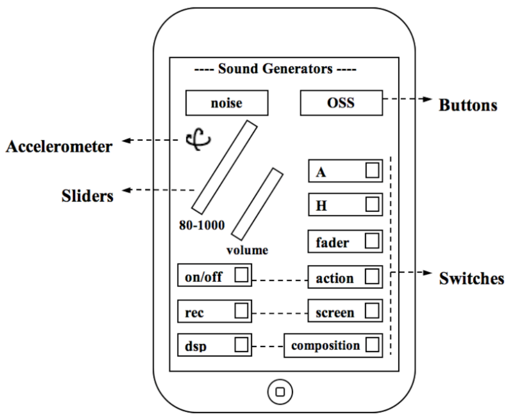

Figure 1. An example of a multi-touch graphical user interface with an accelerometer directly edit in the controller application.

This set of controllers has been selected because they are intuitive and easy to handle. Furthermore, the fact that they

[5] OSCRemote, http://www.nr37.nl/OSCRemote/

have different forms and functions allows one to design a vast repertoire of gesture interaction for music expression. Among the controllers presented, we would like to stress the versatility of the accelerometer, due to the fact that it allows to perform physical gestures of interaction in an open space.

The laptop computer application interface was programmed in Pure Data-extended (Pd-extended) and it is structured in four modules (see Figure 2 below).

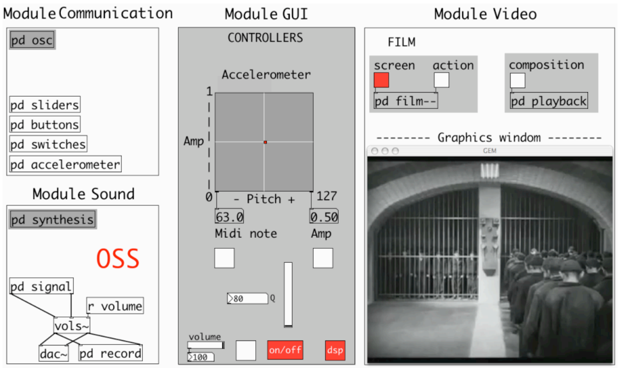

Figure 2. The laptop computer application interface of the MuDI (version 1.0).

Each module executes a specific task. For example, the *Module Communication* carries out the reception, the routing and the mapping data of each controller.

The *Module Sound* is exclusively dedicated to the sound generation and to the recording of the digital musical composition directly onto the laptop computer. The sound generator is constituted by a white noise generator processed by a module of filtering and reverberation and *Our Shepard Sound* (OSS). OSS is a digital sound source based on Shepard Tones. It has the characteristic of being an ambiguous sound in pitch. One could probably identify what OSS musical note is, but it would be very difficult to determine which octave it belongs to. Therefore, this auditory particularity could offer new sonic and musical possibilities of listening.[6]

The *Module GUI* allows visualising the relationship between the gestures performed and the sound generated through the Pd-extended objects called Graphical Users Interfaces or GUI. Each controller of the control application has a GUI object associated to it. Hence, when a controller is manipulated the GUI object associated with it reacts simultaneously allowing both, the performer and the audience visualise the relationship between the gestures performed and the sound generated.

Lastly, the *Module Video* is dedicated to the film's visualisation and to the reproduction of the digital musical composition in synchronisation with the film projected. Start and stop are the only procedures used to control the images of the film.

4. HANDLING

Triggering and scrolling fingers, flexion, extension and internal and external rotation of wrists are the main physical gestures used for operating MuDI.

The relationship between the controllers, the input gestures for interacting, the assigned function of each controller and the final result obtained in terms of the audiovisual perception is outlined in the Figure 3 below.

[6] More information about OSS could be consulted at http://sites.google.com/site/pp2007pt/uk/auditory-illusion

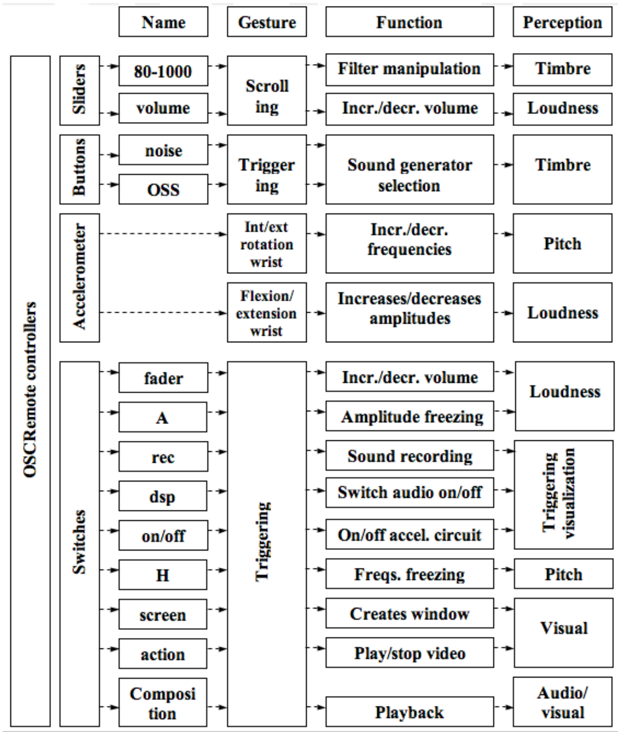

Figure 3. This block diagram represents an overview of the structure and operating mode of the MuDI.

5. CALIBRATION

The process of the calibration can be summarised in the following manner: By placing the motionless handheld device on a table, a reliable central reference point is obtained (red dot of the Figure 4). This reference point represents the calibration of the MuDI's system and corresponds to 0.5 rms of amplitude and to the 63 midi note of pitch.

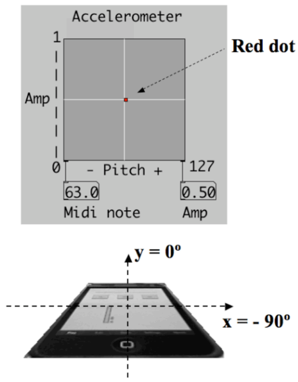

Figure 4. The figure depicts the calibration of the MuDI's system.

The reference point provides both visual and aural feedback to the composer/performer and may be helpful for playing the MuDI.

6. COMPOSITION

In the process of composition the images of the film work as musical cues. The cues could suggest to the composer certain musical gestures. Namely, when start or finish a sound (Length), make a pause or a silence, play slow or fast tempos, play soft or loud (Dynamics), play low or high frequencies (Pitch), choose which sound generator (Timbre), structure the composition in different sections, etc.

On the other hand, the possibility of recording the operating of the laptop computer application interface directly onto the computer (using a video screen capture, e.g. Screenium) provides to the composer/performer a video score. As with any musical score, the obtained video score works as a graphic guide or auxiliary of memory, and subsequently will allow to reproduce (with quite precision) the musical gestures previously composed. These composition aspects can be observed in the video mentioned in the footnote number three of this article.

7. CONCLUSIONS

The multimedia digital instrument (MuDI) belongs to the field of the mobile music performance, a novel musical reality in which everyday common portable media devices are repurposed to become musical controllers and instruments for composing and performing digital music.

MuDI was designed to compose and perform digital music for films which is achieved in real-time. This is the main reason why MuDI permits the viewing of the film images, the recording and the playback of digital musical compositions in synchronisation with the film selected without ever leaving the same programming environment. In this manner, it is thus possible to automatically obtain the digital musical composition performed in real-time in addition to a complete audiovisual file of the work, whilst importing the audio file recorded for the QuickTime Movie file (the file containing the film projected). These procedures allow one to create the soundtrack and the digital music composition of any film.

The introduction of the projection of the laptop computer application interface enabling the performer and the audience to not only see the film images, but also the relationship between the gestures performed and the sound generated. Moreover, the recording of its operating mode allowed to obtain a video score, which proved to be helpful to compose and perform the digital musical compositions mentioned in this article.

The modular approach adopted in the design of the audio engine allows for the sound generator to be reprogrammed and adapted to each individual digital musical composition. If one should desire to obtain a Theremin-like instrument, one merely needs to program, for example, a frequency modulation Pd-extended subpatch and connect it to the “pd synthesis” subpatch (Module Sound). The new sound generator is automatically ready to be used.

The design of the operating mode in question relies upon the need to maintain the simplicity of operation. In addition, the system calibration method which has been implemented has allowed one to overcome the lack of a reliable reference point of the accelerometer. This is a significant point as it is believed that simplicity, reliability and stability improve the confidence of any composer/performer thereby improving musical expression.

The subsequent phase of this area of research will be to continue testing the main features of the MuDI so as to make it even more robust, expressive and user friendly. Further investigation will be carried out in the domain of the sound generators in the “pd synthesis” subpatch (Module Sound). The objective will be to cater for the introduction of additional sound generators capable of going beyond the scope of the white noise and OSS presented in this study. They will be organised so as to allow for an alternate and automatic selection by the performer. This will be achieved as a result of buttons in the handheld controller application. Consequently, the performer will be presented with a wider range of sonic possibilities to play the instrument in addition to being able to create new digital musical compositions in real-time.

8. ACKNOWLEDGMENTS

This work is supported by the Portuguese Foundation for Science and Technology (scholarship SFRH/BD/61599/2009), supervised by António de Sousa Dias. We would also like to thank Luis Gustavo Martins for his suggestions and critical comments.

9. REFERENCES

[1] Essl, G., & Rohs, M. Interactivity for Mobile Music Making. *Organised Sound*, 14(2), 197-207, 2009.

[2] Geiger, G. Pda: Real Time Signal Processing and Processing and Sound Generation on Handheld Devices. *In Proceedings of the International Computer Music Conference* (ICMC) (Singapure), 1-4, 2003. Retrieved 4 October 2010 from http://quod.lib.umich.edu/cgi/t/text/text-idx?c=icmc;idno=bbp2372.2003.*

[3] Geiger, G. Using the Touch Screen as a Controller for Portable Computer Music Instruments. *In Proceedings of the International Conference on New Interfaces for Musical Expressio*n (NIME) (Paris, France), 61-64, 2006. Retrieved 1 November 2010 from http://recherche.ircam.fr/equipes/temps-reel/nime06/proc/nime2006_061.pdf

[4] Tanaka, A. Malleable Mobile Music. *In Adjunct Proceedings of the 6th International Conference on Ubiquitous Computing* (Ubicomp), (Notthingham, England), 2004. Retrieved 4 October 2010 from http://www.ubicomp.org/ubicomp2004/adjunct/demos/tanaka.pdf

[5] Rohs, M., Essl, G., & Roth, M. CaMus: Live Music Performance using Camera Phones and Visual Grid Tracking. *In Proceedings of the 6th International Conference on New Instruments for Music Expression* (NIME), 31-36, 2006. Retrieved 31 January 2011 from http://recherche.ircam.fr/equipes/temps-reel/nime06/proc/nime2006_031.pdf

[6] Essl, G., & Rohs, M. Shamus – a sensor-based integrated mobile phone instrument. *In Proceedings of the International Computer Music Conference* (ICMC) (Copenhagen Denmark), 2007. Retrieved 4 October from http://www.eecs.umich.edu/~gessl/

[7] Dekel, G., & Dekel, A. MoGMI: Mobile Gesture Music Instrument. *In Proceedings of the International Conference on New Interfaces for Musical Expression* (NIME) (Vienna, Austria), 6-10, 2008. Retrieved 4 October 2010 from http://www.mobilemusicworkshop.org/

[8] Wang, G., Essl, G., Penttinen, H. MoPhO: Do Mobile Phones Dreams of Electric Orchestras? *In Proceedings of the International Computer Music Conference* (ICMC) (Belfast, Ireland), 24-29, 2008. Retrieved 10 June 2009 from https://ccrma.stanford.edu/~ge/publish/

[9] Wright, M. Open Sound Control: an enabling technology for musical networking. *Organised Sound,* 10(03), 193, 2005.