Mmmmm: A Multi-modal Mobile Music Mixer

Norma Saiph Savage
UCSB
Computer Science
California, USA
saiph@cs.ucsb.edu

Syed Reza Ali
UCSB
Media Arts and Technology
California, USA
syedali@mat.ucsb.edu

Norma Elva Chavez
UNAM
Computer Engineering
Mexico City, Mexico
normaelvacr@gmail.com

ABSTRACT

This paper presents Mmmmm; a Multimodal Mobile Music Mixer that provides DJs a new interface for mixing music on the Nokia N900 phones. Mmmmm presents a novel way for DJ to become more interactive with their audience and vise versa. The software developed for the N900 mobile phone utilizes the phones built-in accelerometer sensor and Bluetooth audio streaming capabilities to mix and apply effects to music using hand gestures and have the mixed audio stream to Bluetooth speakers, which allows the DJ to move about the environment and get familiarized with their audience, turning the experience of DJing into an interactive and audience engaging process.

Mmmmm is designed so that the DJ can utilize hand gestures and haptic feedback to help them perform the various tasks involved in DJing (mixing, applying effects, and etc). This allows the DJ to focus on the crowd, thus providing the DJ a better intuition of what kind of music or musical mixing style the audience is more likely to enjoy and engage with. Additionally, Mmmmm has an "Ambient Tempo Detection mode in which the phones camera is utilized to detect the amount of movement in the environment and suggest to the DJ the tempo of music that should be played. This mode utilizes frame differencing and pixel change overtime to get a sense of how fast the environment is changing, loosely correlating to how fast the audience is dancing or the lights are flashing in the scene. By determining the ambient tempo of the environment the DJ can get a better sense for the type of music that would fit best for their venue.

Mmmmm helps novice DJs achieve a better music repertoire by allowing them to interact with their audience and receive direct feedback on their performance. The DJ can choose to utilize these modes of interaction and performance or utilize traditional DJ controls using Mmmmms N900 touch screen based graphics user interface.

Keywords

Multi-modal, interaction, music, mixer, mobile, interactive, DJ, smart phones, Nokia, n900, touch screen, accelerometer, phone, audience

1. INTRODUCTION

DJing is an art form that utilizes performance to interact with its audience. Successful DJs can take their audience on a musical journey through their selection of music and how they interweave one track of music to the next. Also equally important is that DJs must play for the crowd, sense how they are feeling and react accordingly. A DJ must identify and understand his audience, i.e. know what interests and triggers it to dance and become animated. A crucial skill DJs must acquire is the ability for them to know their audience from a far. Generally the DJ can be found behind an elevated booth with their hands, eyes, and ears preoccupied on using audio hardware, such as turntables, headphones, mixers, microphones etc, to mix music. While the DJ is managing multiple mixing related tasks, the crowd is enjoying itself; hence the DJ is stuck in solitude, disconnected from his audience. This solitude is also considered one of the most difficult aspects of being a DJ. Despite their upbeat poised attitude, most DJs describe isolation as one of the hardest aspects of their job. In this application, we wished to present an outlet for the conflicts DJs encounter while performing.

2. BODY

2.1 Related Work

Converting mobile devices into musical instruments is a topic that has been tackled in diverse research projects. Specifically, Stanfords mobile phone orchestra [3] aims to use mobile phones as a primary musical instrument for performances. Their research has demonstrated the possibility of using mobile platforms for musical composition, performance, and education. In [3], Wang states: "Mobile music is a new field concerned with the musical interaction in mobile settings, using portable technology." This mentality couple with Wangs work and others who are pushing mobile music forward has contributed to some of the multimodal musical instruments that have become pervasive in society today. These mobile applications such as Ocarina, created by Smule allow users to make the most of a mobile phone as a tool for music creation, composition and performance. Novel interaction methods in this area utilize the embedded sensors in these devices, such as the microphone, accelerometer, and touch screen to provide new interfaces for the areas of research mentioned above.

For example Tanaka, first developed an interface for controlling streaming audio by accelerometer data from an augmented PDA. [7]. Research projects like ShaMus [2] and MoGMI [1], offered gesture-based interactive music making through mobile phones. Much work has been done in using accelerometers and other sensors for mobile interactions. While many of these projects explore the use of the various embedded sensors as a source of input to help cre-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
NIME2010, Sydney, Australia
Copyright 2010, Copyright remains with the author(s).

Figure 1: Mmmmms user interface, the large buttons enable eyes-free gesture based mixing.

Figure 2: This image demonstrates the user manipulating the mixer by using a gesture; the "B" button is used to affect the left track's properties.

ate music, in our project we wished to use these sensors to enhance a DJs ability to mix music through multi-modal interaction. We were particularly interested in the diverse interaction techniques that mobile phones can allow for. Rekimoto introduced the idea of tilting the phone to enable scrolling in an interface, while Williams investigated the use of accelerometers and vibrotactile display for eyes-free text-messaging interactions. Scratchoff presented a platform for experimenting with various types of multimodal feedback while DJing. The findings from Scratchoff suggest that haptic and visual display is interchangeable for the rhythmic gesture tasks a DJ typically utilizes during performance. Moreover, the advances in computing power in laptops and the development of real-time live music composition applications, such as Ableton Live and Traktor Scratch, DJs can perform solely with their laptops without using any additional audio equipment.

As these applications get more robust and the processing power of mobile phones increase, mobile phones should be able to provide the complex functionality of a full fledged virtual DJ application while using their embedded sensors to create a fluid and tangible interface. Recently, a plethora of commercial applications, through which the user can mix tracks on their mobile phone, have been developed and released to the public. The Nokia DJ Mixer, which came out in 2008, allows the user to record songs and remix them with other tracks and get creative with music. It is important to note that the Nokia DJ Mixers interface only allowed input from the user through the smart phones keypad and required the user to watch the screen on all times.

2.2 Design Considerations

When designing a mobile application for DJing or musical performance, there are various factors to consider. First, the application should have the basic core functionality of most desktop virtual DJ applications, that is, it should let users mix at least two tracks of music. Through our background research we found which essential components and functions are essential for a mobile DJ application. These finding were based on using and evaluating popular mainstream virtual DJ applications (Ableton Live, Serato Scratch Live,Traktor Pro, Virtual DJ, and Torq) and recording which functions of the programs were used most. Our findings show the application should be able to play two sound tracks simultaneously with robust sound quality and minimal latency between actions (button presses and other user interactions). The application should be able to cross fade between the two tracks easily and intuitively; the interface should let the user open their music library to select a song to play on a specific track (either track A or track B); the user should be able to control the gain of each track independently so to match volume levels when crossing over to the opposite track; the user should be able to pause and play the track, and seek through the track to a specific cue point in the track.

In addition to evaluating the application functions required for DJing, the requirements for performance with a mobile platform were considered in the interface design. Also, when designing applications for mobile platforms it is important to present the user enough information on the phone screen to allow the performer to get a quick understanding of what is going on in the application. However, if too much information is presented, the interface will be hard to decode and manipulate. Mmmmms interface design takes a minimalist approach. The DJ interface has one main screen so that the user does not have to switch between views, thus allowing them to access all crucial interface elements at all times.

Mmmmms user interface is shown in Figure 1; large buttons "A and "B are in the center of the interface to allow the user to enable gesture control with the need to look at the screen. In the bottom part of the interface we can observe the cross fader and to the sides the volume and song duration, the values of the cross fader as well as the volume can be modified through hand gestures and the user receives feedback of their modification through vibrotactile signals emitted in the phone. In addition, to vibrotactile haptic feedback, the Hildon API was used to pop up notification dialogs when needed (for file loading) and hide them when they were not being used. This is how we managed to fit every element on one screen without overloading the user with information. The interface layout was split into two symmetrical sections, one section for track "A and the other for track "B". Apart from the visual layout, gestures were integrated into the application. Through gestures the user is able to manipulate the interface, changing volume levels of the desired song, cross fading from track A to track B and selecting desired sound effects. These sound effects include: playing the songs in reverse (common music effect used in DJing) by turning the phone upside down, and playing them forward by turning the phone right side up; pitch or tempo shifting a track by placing a finger over the proximity sensor and tilting the one side of the device up (when used in landscape mode) to increase the respective tracks pitch or tempo.

Lastly tapping on the proximity sensor to set the overall tempo of the music. Since our application allows the DJ to perform while taking part in the audience, we provided large buttons in the middle of the interface to allow the DJ to activate gesture control easily and without the need of watching the screen. For each movement that the user carried out, haptic feedback in response was provided. Nokia N900 device offers mobility to the DJ; we use the built-in Bluetooth stereo headphone profile to stream the music to a Bluetooth speaker system, thus allowing the DJ to perform,

Figure 3: Image of a notification as a result of utilizing the Ambient Tempo Detection mode.

Figure 4: This image demonstrates how the public and the DJ interact while using the application.

dance and interact with his audience and environment.

2.3 Technical Details

Mmmmm was developed for the Nokia N900 phone in C/C++ utilizing the gtk, hildon, hildonfm, and gstreamer APIs. The applications interface was developed with gtk and the hildon library. This library enabled us to use special widgets and user interface elements that were specifically designed for the mobile phone. These widgets provided the appropriate sizing for buttons and sliders, so that the interface would be comfortable to use on the touch screen of the N900. Gstreamer was used to create two audio pipelines that run on a separate thread to play two sounds simultaneously and manipulate their properties. This characteristic enables the application to have low latency and be a robust sound player and mixer. The Nokia N900 features a tri-axis accelerometer that can detect acceleration as well as the earths acceleration due to gravity. To carry out the recognition of gestures we used Grant Skinners gesture algorithm. We recorded in a .txt file the recorded movement from a user for a particular movement during a certain time interval. (We recorded the users movement in a .txt file for X seconds.) We did this for thirteen different users and 10 different movements. Each recorded movement from a user had a sequence of X, Y, Z values that were read from the accelerometer. The recorded values were then normalized by determining the "-Y longest edge" of the movement they represented and scaling the rough points "to fit" within a predetermined area. An average for every X, Y, and Z value between thirteen users was computed, and the results were stored in another file, which was denominated "Pattern". The Pattern archive is in fact the training example, with which we compared a new user input with. The users input was compared against the 10 different patterns, i.e. ten different movements. The users input is recognized as a certain pattern when the difference between the users input and the stored pattern input is the lowest in comparison with the other presented gestures.

Although Mmmmm had ten different movements recognized, for the application we only used 5, as to simplify the memory load for the user. These gestures were: rolling, tilting, pointing, hammering and sliding. The user activates the recognition of gestures by pressing either the "A or "B buttons. If the user presses the "A button, the gestures that the user carries out will have an effect on the song that is being played on the left track. If the user presses the "B button, the effects will be seen on the right track.

To detect the "Ambient Tempo, computer vision and the phone video camera are used. A video of the room is taken and processed. Within the processing a classification procedure is made (this will be explained shortly). Using a nearest neighbor approach the video that was taken, is classified and the phone suggests to the DJ the type of music tempo that should be played. For the classification procedure we quantified the amount of movement that existed in the video frame, this was done through frame-differencing, initially the red, green and blue components from each pixel of the first and second image frame are extracted and then the difference of the red, green, and blue values from each of the pixels is obtained. All of the red, green and blue differences are then summed together. For each two frames a value that represents the overall color difference is obtained. We believe this overall color difference maps how fast the audience is dancing or how fast the lights are flashing in a scene. We considered that for a certain range of movement in a scene a certain tempo of music would be adequate; therefore depending on the value that was found a music tempo suggestion is given.

It is important to point out that it is expected that when the DJ initiates the Ambient Tempo detection, he or she keeps the phone stationary without any movement from the DJ, i.e. the camera does not suffer any kind of displacement, as this would affect the calculations. Another aspect that was taken into account for this method, was that because the size of each image frame obtained from the phone camera was significantly large to reduce calculations a down sampling of each of the frames was done.

3. CONCLUSIONS

This paper presented a new multi-modal mobile music mixer, Mmmmm, which is controlled by five different interactive gestures as well as a touch screen; making it one of the first music mixing applications to use embedded sensors and a multi-modal interface on a mobile platform for live DJing. Mmmmm proposes a novel solution for various difficulties encountered while DJing, making the activity much more enjoyable and interactive. Future work will include tests and evaluations of the Mmmmm application for DJing; specifically focusing on the evaluation of the gesture modality to provide an eyes free interface for DJing. Additional types of gestures need to be recognized in order to make a more natural mapping of hand movements into application commands. A user study needs to be conducted to evaluate the performance of the application under typical DJing conditions and its effectiveness as a mobile mixing application. Overall, our application achieved its goal of enhancing the DJing experience with mobile devices, and free the DJs eyes so that he or she can better interact and understand his audience.

3.1 Additional Authors

Rodrigo Savage
UNAM
Computer Engineering
Mexico City, Mexico
rodrigosavage@gmail.com

4. ACKNOWLEDGMENTS

We would like to thank Professor Matthew Turk, who made it possible for us to work with the N900 phone and granted us with his time to brainstorm and discuss ideas and possibilities.

To CONACYT, the Mexican Council of Science and Technology, for funding our studies.

And to Jose Emilio Garcia Acevedo for correcting our text.

5. REFERENCES

[1] A.Dekel and G.Dekel. Mogmi:mobile gesture music instrument. In *Mobile Music Workshop*, 2008.

[2] G.Essl and M.Rohs. Shamus-a sensor based integrated mobile phone instrument. *ICMC*, 3:245–250, 2007.

[3] G. W. Henri Penttinen, Georg Essl. Do mobile phones dream of electric orchestras? *CCRMA*, 1:247, 2007.

[4] R. M.-S. J.Williamson. Shoggle:excitatory multimodal interaction on mobile devices. In *Computer/Human Inteaction*, 2007.

[5] L. Lamport. *LaTeX User's Guide and Document Reference Manual*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1986.

[6] G. E. Nicholas Gillian, Sile O'Modhrain. Scratch-off: A gesture based mobile music game with tactile feedback. *CHI*, 3:234–240, 2009.

[7] H. G. N.Villar. "the colordex dj system: a new live music mixing". *International Conference on new Interfaces for Musical Expresion*, 2:264–267, 2007.

[8] S. R. S.Cho. Dynamics of tilt-based browsing on mobile devices. In *CHI*, 2007.