ZooZBeat: a Gesture-based Mobile Music Studio

Gil Weinberg
Georgia Tech
840 McMillan St.
Atlanta, GA 30332
gilw@gatech.edu

Andrew Beck
Georgia Tech
840 McMillan St.
Atlanta, GA 30332
ab372@gatech.edu

Mark Godfrey
ZooZMobile, Inc.
825 Trowbridge Walk
Atlanta, GA 30350
mark@zoozmobile.com

Abstract

ZooZBeat is a gesture-based mobile music studio. It is designed to provide users with expressive and creative access to music making on the go. ZooZBeat users shake the phone or tap the screen to enter notes. The result is quantized, mapped onto a musical scale, and looped. Users can then use tilt and shake movements to manipulate and share their creation in a group. Emphasis is placed on finding intuitive metaphors for mobile music creation and maintaining a balance between control and ease-of-use that allows non-musicians to begin creating music with the application immediately.

Keywords: mobile music, gestural control

1. Introduction

ZooZBeat is a gesture-based mobile musical studio designed to unlock musical expression and creative potential for novices and experts alike. It currently runs on iPhone, iPod Touch, and the Nokia N95 smartphone. Unlike musical rhythm games such as Guitar Hero and Rock Band that can be seen as "music-influenced games" (traditional eye-hand coordination and speed challenges are mapped to musical outcome), ZooZBeat focuses on providing immediate self-expression based on musical and game theories. Through a set of easily learned, natural gestures, ZooZBeat users actively enter musical material that is processed to fit the current musical context and entered into a looping sequencer. The loop-based interaction is forgiving of mistakes while providing structure. Here, users can perform additional gestures to manipulate and share their creation in a group. A key goal in designing ZooZBeat is to provide immediate engagement and self-expression for novice players, while providing a wide room for practice, improvement, and virtuosity that would engage more experienced musicians as well.

2. Related Work

Advances in consumer mobile phones have facilitated massively distributed applications for music generation. Researchers in the field of New Interfaces for Musical Expression (NIME) have developed mobile music applications that, in comparison to older computer-based musical systems, can better support the immediate, self-contained, and mobile nature of many acoustic musical instruments. Embedded accelerometers in mobile devices, for example, allowed researchers to experiment with gestural control of streaming audio [1] or self-contained audio synthesis [2]. Other input devices that have been used for musical and audio driver application include touch screens [3,4], embedded cameras [5], and GPS [6, 7]. Recently, several music-making applications for mobile devices have emerged, from a breath-controlled wind instrument [8] to full-featured drum machines [9, 10] and synthesizers [11]. Most of these experiments, however, did not aim to offer an integrated system that allowed novices and seasoned musicians, to generate, edit, sequence, and share their original compositions.

3. Application Structure – Songs and Gestures

At the core of ZooZBeat is the song - a collection of MIDI instrument tracks, background loops, and an organization of loops and related scales into a complete song structure. The format of the song allows full length, multi-sectional pieces to be condensed into short and manageable periods of time. Sections that have nearly identical backing tracks are reduced into a single background loop and attached to multiple parts of the song. For instance, if the chorus repeats five times in a song, only a single loop is needed to represent all occurrences. Setting the act of music making in the context of a pre-organized song makes the process less intimidating to inexperienced users, allowing for uninhibited immediate engagement.

For musicians, a set of complex muscle memory mappings defines the sensation of playing music. Subtle variations in finger and arm movements create the feeling of playing a guitar, piano, or drum. For the non-musician, musical gestures are based more on the experience of watching musicians play rather than their own muscle memory [12]. Our main design goal therefore was to find intuitive mappings between gestures and a loop-based music sequencer, ones that users, both musicians and non-musicians, can immediately recognize and interact with.

3.1 Shaking

The most basic note entry mechanism is that of shaking the phone. This gesture is similar to hitting a drum or striking a piano note. Onset detection analysis is run on input from a built-in accelerometer (if no accelerometer is present, optical flow based on input from the phone camera is used). The energy level of the onset is determined and mapped to the new note's pitch: the harder the energy of a shake, the higher-pitched the sound. In this way, melodic contours are created by adjusting how hard the phone is shaken. This shaking gesture has proven to be a widely understood gesture, since many instruments rely on some sort of striking gesture. Most people have some experience playing drums and can easily relate to this gesture.

3.2 Tilting

One gesture that is commonly associated with musical expression is tilting an instrument while playing. Classical musicians often move their violin or flute around during particularly emotional sections. Rock guitarists often tilt their guitars up during intense solos, communicating an expressive and emotional impression to viewers [13]. In ZooZBeat, users can experiment with such gestures by tilting their phone up and down and rolling it left and right. The tilting gesture transforms previously recorded material, subdividing note durations in the stored sequence. The sequenced notes are subdivided into 2 (by tilting upward) or 3 (by tilting downward) equal parts. The extra notes are considered ornamentations of the original melodic line, and their pitches are resolved by a set of deterministic rules governed by surrounding pitches.

3.3 Tapping

Another method of note entry in ZooZBeat is tapping notes on a keypad. On most phones, this involves pressing physical keys, holding them for the duration of the note, and releasing when finished. On touch-screen devices without a keypad, users can tap the screen instead. Lower notes are obtained by tapping lower keys (or tapping lower on the screen), and higher notes by pressing higher keys (or tapping higher on the screen). This method allows users to enter notes with a finer grain of detail than by shaking the phone. A primary mechanism of interaction with a phone, tapping the keys is a simple gesture for all users.

3.4 Tossing

For multi-player games where only one player is interacting with the system at any given time, the tossing gesture is used to pass the music to other phones in the group. This feature simulates jam session scenarios such as in drum circles or in jazz improvisation, where the lead is passed around among the musicians. Here, users choose the phone to which they will transfer the musical control and "throw" the music in that direction.

4. Modes of interaction

4.1 Free Individual Play

Free play is the most straightforward performance mode, intended for boundless musical expression without goals or complex interactions. When the player starts a song a background loop is playing, and a selection of instruments are displayed on-screen. In addition to three percussive and three melodic instruments, users can also choose the microphone track, where they can record their voice (or any sound other input) and add it to the song. Each background loop is tied to a quantization level and scale so that the player can add new tracks in a relevant tonal and rhythmic context. The user selects an instrument by choosing it from a menu and can add notes using shaking, tilting, and tapping gestures (as described in Section 3).

4.2 "Hot Potato" group play

In this mode players pass control of the sequencer through the "tossing" gesture described in Section 3.4. Each player is assigned a track, which can be "tossed" to the next player when ready. The receiving player can then play and add a new track, while each original player can still manipulate the track he or she sent using tilting gestures. In this way, participants are offered a sense of ownership of their track, allowing for continuous engagement in the composition. This approach reduces potential confusion caused by too many simultaneous musical inputs.

Figure 1. Three players engage in "Hot Potato" group play

4.3 "Copycat" group play

Like in "Hot Potato" mode, "Copycat" mode begins with each player assigned a track. When a player plays notes into a track, the notes are tossed to the next player, who is challenged to play back in another track what he sees and hears in a set amount of time. The receiving player can then create a new track to challenge another player. This mode is similar a multi-player game as points are given based on how well a user can replay the last player's melody.

Figure 2. Screenshot of ZooZBeat for iPhone

5. Software

A shared code base was written in C to maximize support of as many next generation phone platforms as possible. The only requirements to run this code base are file input/output, an audio output stream, and the ability to run compiled C code. The advantages of a shared code base include the ability to synchronize among several platforms and the ability for multiple platforms to share common media files. The code base, inspired by popular music sequencers, organizes a song into samples, instruments, background audio loops, and sequences.

5.1 Audio Engine

A sample is a piece of audio tagged with information such as sample rate, loop positions, base frequency (in MIDI note number) and current state. Each sample is constantly passed to the output stream. If the sample is marked as playing, then the appropriate audio bytes are fed to the output stream. This method allows for thread safety when dealing with platforms such as the iPhone, where samples are triggered by the event thread and the audio bytes are outputted by the audio thread.

An instrument is a collection of samples, laid out in the same way as popular software samplers, each sample given a MIDI-note range. Sequences are represented by a collection of instrument tracks, MIDI note numbers recorded, a scale, and a collection of background audio loops. Notes are entered asynchronously into the sequence and instruments are triggered when a note is encountered. In this way, the behavior of popular music sequencers is emulated.

Since ZooZBeat is running on a mobile phone, a built-in microphone is always present, allowing for sampling of live audio, typically of the user's voice. One instrument track is dedicated to playing back these recordings. After recording, the sequencer divides the new sample into several sections and considers each a note, with higher "pitch" mapped to later occurring sections of the sample. Users can then produce scratching or stuttering effects on the recording by entering and manipulating notes on this track.

5.2 Wireless Communication

Some phones support wireless protocols for data transport. In these phones, we can implement synced multi-player modes. On the Nokia N95, for example, we use the Bluetooth RFCOMM protocol to create a master-slave relationship between up to eight phones. On the iPhone, we use a standard IEEE 802.11 wireless connection to synchronize multiple devices. The exact protocol used for synchronizing tempo and data is transport agnostic and is designed to be robust in the presence of dropped packets.

The most challenging data to synchronize is the beat alignment. Even when two phones start at exactly the same time, slight changes in clock speed between the phones can put them out of phase by as much as a 16th note after a few bars. To synchronize the timing of two phones, we need to estimate the time it takes for a packet to move from one phone to the next. To do this, we send a number of messages back and forth and measure the amount of time from sending each message to receiving the response. Once the system is confident that there is low variance between most of the estimated timings, it sends a final sync message, signaling to the other phone that the next beat is a specified amount of time away, taking into account the estimated send time. If the internal timing on the receiving phone is drastically different from the timing sent by the master phone, the receiving phone will adjust itself only slightly to prevent large jumps in sequencer timing.

Global sequencer information is kept on the master phone. Each phone retains an individual copy of the current sequencer and is kept up to date through messages from the master phone. Updated sequencer information is sent back to the master.

5.3 User Interface

In line with our desire to allow for natural gestural control over musical events and multi-user interaction, ZooZBeat's user interface was designed to welcome users inexperienced with technology-aided music production while streamlined and powerful enough so as not to hinder musical expression. As seen in Figure 2, a horizontal staff-like representation of a track's content is displayed at the bottom of the screen, similar to many computer-based sequencers. To avoid confusion and clutter, only the current track's notes are displayed. On the iPhone, this section of the screen doubles as a gestural input: a touch on this section from the user generates notes as described in Section 3.3. Additionally, tilting the phone on its side causes it to enter a landscape note-entry mode where the

touchable section is larger and a grid is superimposed, allowing for finer control over note pitch and location.

Users switch tracks (and thus instruments) by selecting an icon from the sides of the screen. On the iPhone, this is accomplished by a simple touch of the icon, while the Nokia requires scrolling through each with the directional pad. These instrument icons also serve as a visual cue to the source of heard sounds as they vibrate when a note is played on its track. This also gives the application a whimsical sense of "life" as the icons appear to be dancing.

6. Preliminary User Feedback

The response to ZooZBeat has been quite varied. Many have been delighted at the ease by which fairly complicated musical material is generated, while others find the resulting music trite and the quantization algorithms heavy-handed. Musicians and non-musicians alike find the application enjoyable and at times surprisingly inspiring. Typically, musicians experienced with computer-based music production tend to look for more advanced features that were intentionally omitted for simplicity's sake, such as volume and tempo control.

Reviews of the gestural controls have been generally positive. TechCrunch considers them a "perfect use of the built-in accelerometers… you can pick an instrument and simply tap, shake or tilt to create your own masterpiece" [14].

Many users complain that there is no way to export their creations from the phone to their home computer for later listening or remixing. Wired's Eliot Van Buskirk, after being "impressed" and "thoroughly enjoying" creating loops, "really wanted to upload [his] creations" [15]. To address this, a web component of ZooZBeat, called ZooZme, is under development that will allow users to upload their songs, and from there, listen and share their creations with other users.

Interestingly, another common complaint was that the application seemed "pointless", as there is no goal or finale. Familiar with games like Guitar Hero, typical consumers of such applications seem to tend toward more structured play, where their actions are guided by more than a desire to produce satisfying music.

7. Future Work

Currently under development are social interaction features that will focus on enabling users to play with ZooZBeat in groups. Most simply, this involves sharing finished songs with your friends, while more ambitious goals include taking the ad-hoc local area wireless communication protocol discussed earlier and adapting it to a client-server architecture so that people across continents can jam together. We are also exploring more play modes similar to the described "Copycat" multi-player game, where users are driven by competition while also encouraged to be creative and musically expressive.

Enhancements to the audio engine are also under development that will allow for optimized audio effects whose parameters are controllable in real-time. These will not only satiate musicians eager for more production power but will also allow non-musicians the opportunity to intuitively explore another dimension of music: timbre.

References

[1] A. Tanaka. "Mobile Music Making," in *Proc. of the Int'l Conf. on New Interfaces for Musical Expression (NIME),* pp. 154–156, Hamamatsu, Japan, 2004.

[2] G. Essl and M. Rohs. "Mobile STK for Symbian OS," in *Proc. Int'l Computer Music Conf. (ICMC)*, pp. 278–281, New Orleans, USA, 2006.

[3] G. Geiger. "PDa: Real Time Signal Processing and Sound Generation on Handheld Devices," in *Proc. of the Int'l Computer Music Conf. (ICMC)*, Singapore, 2003.

[4] G. Geiger. "Using the Touch Screen as a Controller for Portable Computer Music Instruments," in *Proc. of the Int'l Conf. on New Interfaces for Musical Expression (NIME)*, pp. 61 – 64, Paris, France, 2006.

[5] M. Rohs, G. Essl, and M. Roth. "CaMus: Live Music Performance using Camera Phones and Visual Grid Tracking," in *Proc. of the Int'l Conf. on New Instruments for Musical Expression (NIME)*, pp. 31–36, Paris, France, 2006.

[6] S. Strachan, P. Eslambolchilar, R. Murray-Smith, S. Hughes, and S. O'Modhrain. "GpsTunes: Controlling Navigation via Audio Feedback," in *Proc. of the Int'l Conf. on Human Computer Interaction with Mobile Devices & Services*, pp. 275-278, Salzburg, Austria, 2005.

[7] A. Tanaka, G. Valadon, and C. Berger. "Social Mobile Music Navigation using the Compass," in *Proc. of the Int'l Mobile Music Workshop*, Amsterdam, 2007.

[8] "Smule: Ocarina," [Web site] 2008, [2009 Jan 29], Available: http://ocarina.smule.com/

[9] "Intua: BeatMaker," [Web site] 2008, [2009 Jan 29], Available: http://www.intua.net/products.html

[10] "Izotope: iDrum," [Web site] 2008, [2009 Jan 29], Available: http://www.izotope.com/products/audio/idrum/iphone/

[11] "Noise.io," [Web site] 2008, [2009 Jan 29], Available: http://noise.io/

[12] M. Rodger, J. Issartel, and S. O'Modhrain. "Performer as perceiver: perceiver as performer," in *Proc. of the Int'l Conf. on Enactive Interfaces (ENACTIVE/07), Grenoble, France, 2007.*

[13] C. Cadoz and M. M. Wanderley. "Gesture – Music," in *Trends in Gestural Control of Music*, M. M. Wanderley and M. Battier, Eds. Paris: IRCAM, 2000, pp. 71-94.

[14] "ZooZBeat Turns Your Phone Into a Music Studio," [Web site] 2008, [2009 Jan 29], Available: http://www.techcrunch.com/2008/11/14/zoozbeat-turns-your-phone-into-a-music-studio/

[15] "ZooZBeat Turns iPhone into Beat Factory for Three Bucks," [Web site] 2008, [2009 Jan 29], Available: http://blog.wired.com/business/2008/12/zoozbeat-turns.html