Tangible Beats: a collaborative user interface for creating musical beats

Alexandre M. F. de Sousa and Carlos H. Morimoto

Laboratory of Technologies for Interaction
Department of Computer Science
University of Sao Paulo, Brazil
{alemart, hitoshi}@ime.usp.br

Abstract

Tangible User Interfaces (TUIs) allow interaction with digital content through physical artifacts, enabling a smooth bridge between the digital and the physical worlds. In an entertainment context, such interfaces offer rich expressive capabilities for experimental works. This paper introduces Tangible Beats: a user interface that, in a collaborative fashion, enables the creation of musical beats by manipulating colored tokens on a flat surface. A prototype implementation, composed of a Kinect camera, a video projector, speakers and a laptop is presented. The Kinect is used to estimate the position of the tokens and the projector provides visual feedback to help inexpert music makers compose their own beats by changing the notes and the instruments.

Keywords: tangible user interfaces, interactive art, collaborative music interface, exploratory design

1 Introduction

A music box is a mechanical music instrument. A comb-like metal plate with tuned teeth is sounded by pins spread over the surface of a revolving cylinder (see Figure 1). This straighforward mechanism for producing sounds inspires novel ways to help inexpert music makers compose music in a real-time collaborative manner.

Human beings have developed complex skills for interacting with the physical environment. Grasping objects is prevalent in our everyday lives. Nowadays, however, most of these abilities are not explored when interacting with the digital realm. In Ishii and Ullmer's vision of Tangible User Interfaces (TUIs), digital content is coupled with physical objects, enabling users to "grasp and manipulate" information [Ishii and Ullmer 1997]. In the fields of interactive arts and games, TUIs give rise to many creative possibilities. In particular, one can develop a TUI that builds upon the concept of the music box, introducing elements of entertainment, interactivity and collaboration.

This paper presents Tangible Beats: a novel user interface that enables the creation of musical beats by manipulating colored tokens on a flat surface. Tangible Beats provides both tactile and visual representations of the beats, allowing users to intuitively "grasp" the abstract essence of the music (see Figure 2). By manipulating the tokens, novice players make and experience new musical patterns collaboratively and in real-time.

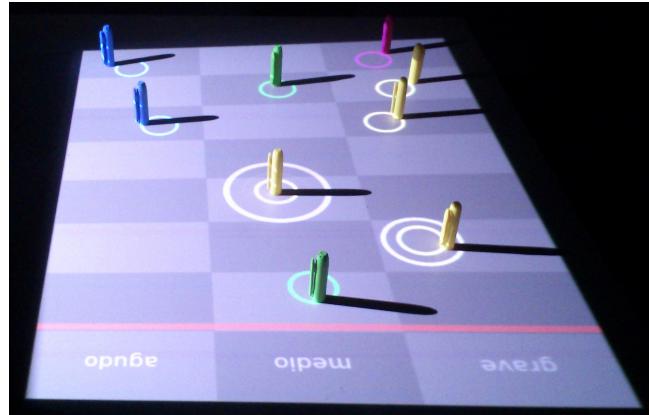**Figure 1:** *The internals of a music box [Eugster 2010]*

Figure 2: *Tangible Beats*

2 Related Work

Exploratory design in the field of computer-assisted musical interfaces has been an active theme of research over the last few years. Using the inexpensive Nintendo Wiimote sensor, Fornari has proposed a gesture-based interface that leverages human body movements for musical expression [Fornari 2010]. The accelerometer data is mapped to computer models of sound synthesis. The popular reacTable [Jordà et al. 2007] has introduced an innovative design based on a Tangible User Interface that enables performers to compose their complex rhythms collaboratively in real-time. Having reached massive popularity outside the academic world, it eventually became a commercial product.

Focusing on inexpert music makers, The Music Table [Berry et al. 2003] presents a way of composing musical patterns by arranging cards on a tabletop. Each card represents a note. By moving the cards, a user can modify the pitch of the notes, as well as their timing in a looping timeline. A camera tracks the positions of the cards. Real-time feedback is provided in the form of audio and on-screen computer-generated images. A special instrument card allows the user to change the instrument of the notes being played. Tilting the card cycles through a 'palette' of pre-programmed instruments. A contemporary Japanese sensibility inspires the visual aesthetics: 3D creatures are rendered on top of the cards in an augmented-reality fashion. A large screen placed near the tabletop provides visual feedback. Although the system brings about a physical representation of the music, the visual feedback is detached from the interaction space. Using different musical instruments can be cumbersome, requiring many steps to modify each note or set of notes.

As a Tangible User Interface, BeatBearing [Bennett and O'Modhrain 2008] exploits users' "natural" ability at manipulating physical objects, making use of physical constraints and affordances. In order to provide the experience of "flow", BeatBearing brings an interface that appears "self-contained". The user is presented with a physical control surface, a screen and an audio source that share the same interaction space. On top of a horizontally-placed flat-screen monitor, there are custom-made electronics that take the shape of a grid. The musical rhythm is composed by placing the balls on holes spread over the grid. Modifying the position of the ball on the x and y axis change the timing and the instrument of the note, respectively. A moving line is displayed on the monitor. As it sweeps past a ball, the appropriate note is played. BeatBearing allows the "hands-on" manipulation of the rhythm, although changing the pitch of a note is not possible.

3 Design

The design goal of this work is to allow inexperienced users play with music collaboratively and intuitively. As a case of exploratory design, Tangible Beats builds upon the concept of the music box. Its simple mechanism can be easily understood by people without any musical background, making it the inspirational case of our choice. Due to the physical constraints and affordances of the tokens, multi-user collaboration is a natural consequence of the design. Users simply have to move the tokens across the tabletop to experience the resulting musical beats.

Tangible Beats features:

- a rectangular grid of 8×3 cells that is projected over the tabletop;

- 4 colored tokens of 4 different colors (yellow, blue, green, magenta), totalizing 16 tokens;

- a vertical red scanning line that moves horizontally, repeatedly disappearing on one side of the grid and reappearing on the other side.

As shown in Figure 3, users compose their beats interactively and collaboratively by placing the tokens on the cells of the grid. The physical model of the music works as follows:

- each physical token represents a musical note;

- tokens of different colors are mapped to different instruments;

    - The yellow, blue, green and magenta tokens are mapped to hi-hat, drum, bass and "beep", respectively.

- each cell of the grid can hold a musical note;

- the horizontal axis of the grid represents time and the vertical axis represents the pitch of a note.

The closer a token is to the "bottom" of the grid, the lower the pitch. Conversely, tokens placed around the "top" of the grid are mapped to a higher pitch. Whenever the scanning line moves past a token, its corresponding note is played (in that moment, a transparent halo is displayed temporarily for additional visual feedback). Combining the dimensions of time and pitch allows an easy-to-learn mechanism to compose of the beats. Since the tokens can be placed anywhere in a continuous physical space, the grid gives a discrete characteristic to the sounds: it is similar to a simple compass. Figures 4a, 4b, 4c and 4d show different beats that can be created with the system.

The faster the scanning speed, the more beats per second will be experienced by the user. Users can modify the scanning speed using a two-handed gesture. Whenever a finger is placed on the grid (call it finger A), a speed modification mode is activated. Touching the grid with another finger (say, finger B) allows the modification of the speed: moving B in the direction of the movement of the scanning line increases its speed, and moving B in the opposite direction decreases its speed. By withdrawing finger A from the tabletop, the user deactivates the speed modification mode.

Table 1 summarizes the connection between the features provided by the interface and the auditory feedback.

Auditory feedback	Interface feature
Instrument	Tokens of different colors
Location of the beats in time	Horizontal position of the tokens
Pitch	Vertical position of the tokens
Tempo	Two-handed gesture

Table 1: *How interface features map to auditory feedback*

Colored halos are projected over the tokens as a visual feedback that the system has recognized the physical objects. Although the scanning line is a virtual entity, a sound is played whenever it touches a token. The tabletop becomes a bidirectional interaction device, as the user communicates its intentions and receives the corresponding visual feedback on the table. Input and output of data happens in the same interactive space.
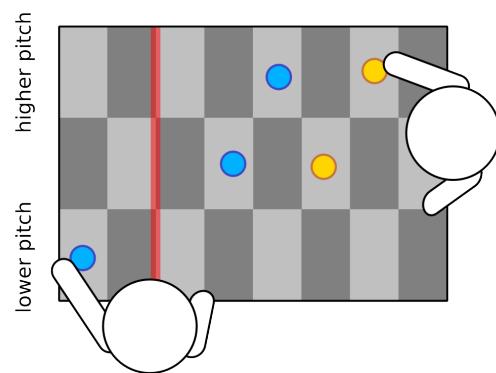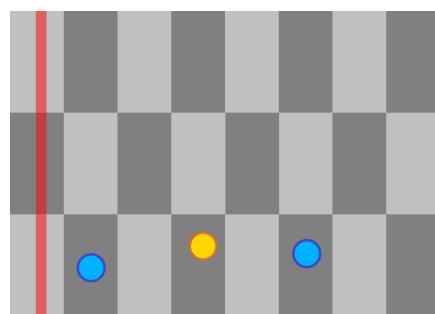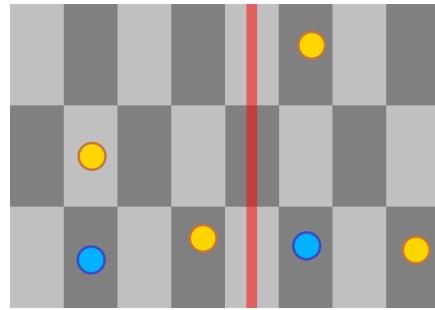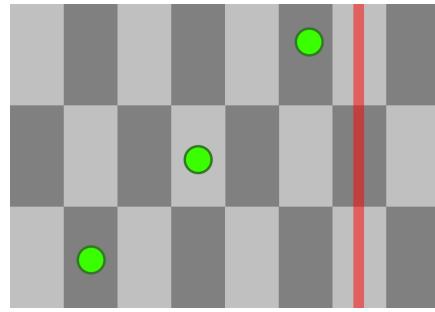

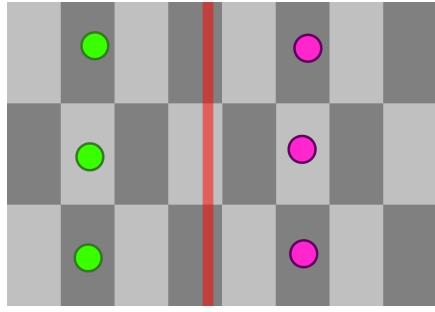
Figure 3: *Collaborative composing*

(a) *A simple beat*

(b) *A more sophisticated beat*

(c) *Bass: from lower to higher pitch*

(d) *Multiple notes being played at the same time*

Figure 4: *Sample beats*

4   Implementation

The prototype implementation, shown in Figure 5, is composed by three subsystems: Touch Detection, Color Recognition and Music Player. The Touch Detection module, responsible for converting a flat surface into a touch sensitive region, determines the position of the tokens placed on the tabletop. The Color Recognition subsystem is responsible for finding the colors of the tokens. Once the position and the colors of the tokens have been determined, the Music Player provides both visual and auditory feedback. The hardware setup includes a laptop computer, a Microsoft Kinect, speakers and a small video projector attached to a tripod.

4.1   Touch Detection

The Microsoft Kinect device includes a depth sensor and a color camera. These resources make it a feasible tool for detecting and tracking the colored tokens spread over a tabletop. Using a Kinect, Wilson has described a technique that detects touching objects on a non-instrumented tabletop [Wilson 2010]. Although the results are not as accurate as capacitive touch screen technologies, the approach is appropriate for various applications, with the added benefit that the form of the objects may also be exploited. Xiao, Harrison and Hudson have used a similar technique in WorldKit, an interactive system that augments everyday surfaces [Xiao et al. 2013]. This work combined the techniques presented on Wilson and on Xiao, Harrison and Hudson, with some modifications.

A depth image is transformed into a set of blobs corresponding to the physical tokens that are placed over the tabletop. This is accomplished by performing a sequence of image processing operations: background calibration (with an empty table), determining of a noise profile, background subtraction, thresholding and additional filtering using erosion and dilation operations. The end result will consist of blobs on a binary image. Each blob is identified using connected component analysis. The center of mass of the resulting components indicate the position of the tokens. Their position $P_k = (x_k, y_k)$, in depth image coordinates, are transformed to a $[0, 1] \times [0, 1]$ space using a homography matrix, computed with a standard four-point calibration procedure [Bradski and Kaehler 2008].

4.2   Color Recognition

In depth image coordinates, for each blob detected in the previous phase, we map all pixels in a sphere of fixed radius r centered in each T_j to the color image provided by the Kinect. Then, the colors of all those pixels are converted to CIELAB space. After filtering out their luminosity component, we are left out with a cluster C of pairs of numbers representing the chromacity of each physical token. Based on the observation that the chromacity cluster is approximately elliptical in shape, the color recognition subsystem is built upon Lee and Yoo's elliptical boundary model [Lee and Yoo 2002], originally proposed as a binary skin classification model. Given a set of k known colors, $\{j_1, j_2, \ldots, j_k\}$, our method extends their work as follows: cluster C is classified by picking j that maximizes the density

$$d_j = \frac{|\{ p_i \in C \mid (p_i - \phi_j)^t \Lambda_j^{-1} (p_i - \phi_j) < \theta \}|}{|v_j|} .$$

θ is an empirically specified parameter and $\phi_j \in \mathbb{R}^2$ and $\Lambda_j \in \mathbb{R}^{2 \times 2}$ are the model parameters computed according to [Lee and Yoo 2002]. The signed volume v_j is the determinant $|\Lambda_j|$.

4.3   Music Player

The modules described previously cover the computer vision component of this work. They provide a set of triples (x, y, c) that indicate the position and the color of the physical tokens. This data is converted to visual and auditory output by the Music Player. The Music Player has been developed using openFrameworks, a multi-platform C++ toolkit that provides vast resources for interactive artists [Marinho et al. 2008]. The 8×3 grid is rendered using
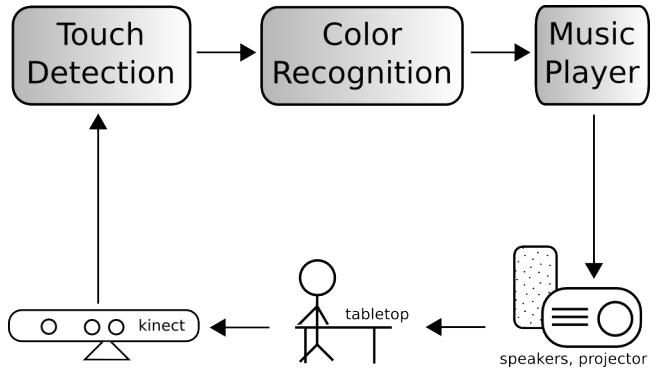

Figure 5: *System overview*

neutral colors, ensuring that the hue of the colored tokens is preserved in the presence of a video projector.

5   Tests and Results

We have presented Tangible Beats twice at an art exhibition in our university. During the exhibition, guests would freely play with the system. Users ranged from young children to mature people. The public was mostly composed of university students, visiting high schoolers and local employees. The prototype setup is displayed in Figure 6.
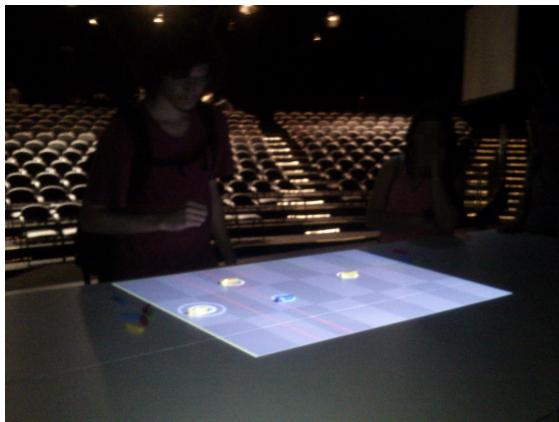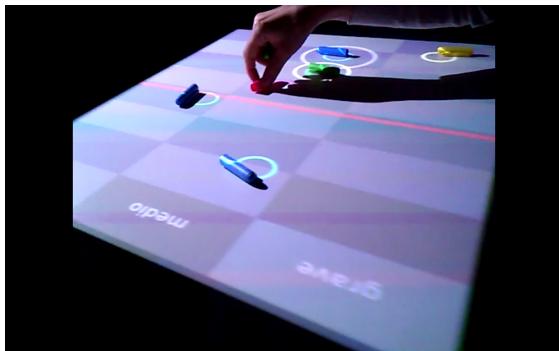

Figure 6: *The prototype setup*

In our tests, we found out that people would generally "discover" a beat interactively (see Figures 7a, 7b and 7c). They would manipulate the tokens until they could hear something that "made sense" to them. Although we had given users a basic explanation of what the system did, we observed that most of them would rather figure out by themselves how to use it (probably due to the ludic nature of the prototype). When in doubt, users asked whatever they wanted to know. Collaboration happened spontaneously, particularly among couples and groups of friends.

User feedback tended to be quite enthusiastic: "I got addicted to it, can't stop playing! It's highly interactive!", "fun, amazing, new way of playing an instrument", "I liked it because I created a sound that could be repetitive, that seemed like music". Such feedback suggests that the design introduced by this work is well suited to an entertainment context.

Although the prototype works well in a controlled laboratory setting, there were a few issues during the exhibition. The tabletop was placed in a theater stage with shaky ground. That introduced a bit of noise from time to time. Additionally, due to adverse lighting conditions, our color recognition module failed at times. Since it could not pick up skin color correctly, there were problems testing the two-handed gesture for controlling the speed of the scanning

(a) *Tangible Beats at the art exhibition*

(b) *A user manipulating the tokens*

(c) *A beat codified using the physical tokens*

Figure 7: *The working prototype*

Acknowledgements

References

BENNETT, P., AND O'MODHRAIN, S. 2008. The beatbearing: a tangible rhythm sequencer. In *Proc. of NordiCHI*, vol. 2008.

BERRY, R., MAKINO, M., HIKAWA, N., AND SUZUKI, M. 2003. The augmented composer project: The music table. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, IEEE Computer Society, 338.

BRADSKI, G., AND KAEHLER, A. 2008. *Learning OpenCV: Computer vision with the OpenCV library*. O'reilly.

EUGSTER, S. A., 2010. File:Musical box open.jpg - Wikimedia Commons. [Online; accessed 22-July-2014]. Available at: http://commons.wikimedia.org/wiki/File:Musical_box_open.jpg.

FORNARI, J. 2010. Gestural controlled virtual percussion instruments. In *IX Brazilian Symposium on Games and Digital Entertainment (SBGames 2010), Florianpolis, SC–Brazil*.

ISHII, H., AND ULLMER, B. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, ACM, 234–241.

JORDÀ, S., GEIGER, G., ALONSO, M., AND KALTENBRUNNER, M. 2007. The reactable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, ACM, 139–146.

LEE, J.-Y., AND YOO, S. I. 2002. An elliptical boundary model for skin color detection. In *Proc. of the 2002 International Conference on Imaging Science, Systems, and Technology*, Citeseer.

MARINHO, B. N. S. F. C., BERGAMO, M. L., AND MOTA, R. R. 2008. Programação para artistas. In *VII Brazilian Symposium on Games and Digital Entertainment (SBGames 2008), Belo Horizonte, MG–Brazil*.

WILSON, A. D. 2010. Using a depth camera as a touch sensor. In *ACM international conference on interactive tabletops and surfaces*, ACM, 69–72.

XIAO, R., HARRISON, C., AND HUDSON, S. E. 2013. Worldkit: rapid and easy creation of ad-hoc interactive applications on everyday surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 879–888.

line. According to a local musician, such a control would be a "huge feature for creative expression". Possible workarounds for this issue include: retrain the color models in adverse lighting conditions, introduce a virtual scrollbar on the tabletop.

6　Conclusion

We have presented Tangible Beats: a collaborative user interface that allows the creation of musical beats in real-time by manipulating colored tokens on a tabletop. The visual and tactile representation of the beats enables inexpert music makers to intuitively create their own musical patterns. A working prototype, built with a Kinect, a video projector, speakers and a laptop, was presented at a local art exhibition. The enthusiastic reaction of most users showed that this design experiment introduces a collaborative, easy to learn and fun way of creating musical beats, particularly for novice players. One possibility for a future work is the use of special markers as a "physical memory" to store user-created beats. Playing in the background, those stored beats could be combined with whatever the user composes with the colored tokens.