

INTRODUCING AUDIO D-TOUCH: A TANGIBLE USER INTERFACE FOR MUSIC COMPOSITION AND PERFORMANCE

E. Costanza
Media Engineering Group
Dept. Of Electronics
The University of York
York, UK
e.costanza@ieee.org

S.B. Shelley
Media Engineering Group
Dept. of Electronics
The University of York
York, UK
simonbshelley@yahoo.co.uk

J. Robinson
Media Engineering Group
Dept. of Electronics
The University of York
York, UK
jar11@ohm.york.ac.uk

ABSTRACT

"Audio d-touch" uses a consumer-grade web camera and customizable block objects to provide an interactive tangible interface for a variety of time based musical tasks such as sequencing, drum editing and collaborative composition. Three instruments are presented here. Future applications of the interface are also considered.

1. BACKGROUND

Tangible User Interfaces (TUIs) are a recent field of research in human-computer interfaces. Physical objects are used for the control and representation of digital information [1, 2, 3]. Where in a Graphical User Interface (GUI), users interact with virtual objects represented on a screen through mice and keyboard, in a TUI the users physically place and move objects in the real space to achieve the same results. Generally each physical object represents a particular piece of digital information or part of a virtual model. The computer output is usually presented in the same physical environment to reinforce the perceptual link between the physical and virtual objects. Grasping an object becomes then analogous to grasping the corresponding piece of digital information. Consequently, the interaction tends to be more direct and the user is allowed to take advantage of his advanced and multisensory perception of real items (compared to the purely visual perception through a screen – a two-dimensional window on a virtual world) [4].

A number of researchers have already used tangible user interfaces for musical applications [5-8]. For example, the Augmented Groove [7] allows modulation of pre-recorded samples, and Block Jam [6] allows selection of one of a number of pre-recorded samples on each block in a TUI. However, neither exploits the flexibility of independently movable objects to give broad functionality. Moreover, their use of expensive and sometimes fragile technology – such as sensors and displays – embedded in the tangible interactors, mitigates against their widespread use.

In audio d-touch the user can create patterns and beats, rather than adjusting preset ones. In general we use direct mapping of physical quantities to musical parameters (such as timbre and frequency), resulting in simpler interaction. This simplicity, however, does not prevent the interface from being very flexible and allows advanced users to research and create rich and complex sound textures. We stress the use of analogy both in terms of similarity to traditional music notation and in terms of high

definition (virtually continuous) mapping between input and output.

The system tracks the position of interactive objects with a web-cam, by means of a robust image fiducial recognition algorithm. In order to make an object recognisable and interactive, it has to be marked with the fiducial pattern. Technical details can be found in [4, 9]. This technology is significantly less expensive to that required for comparable systems. As a consequence our instruments are targeted at musicians and museums as well as home users and schools.

We report three musical applications based on a TUI and discuss the novelty of our approach.

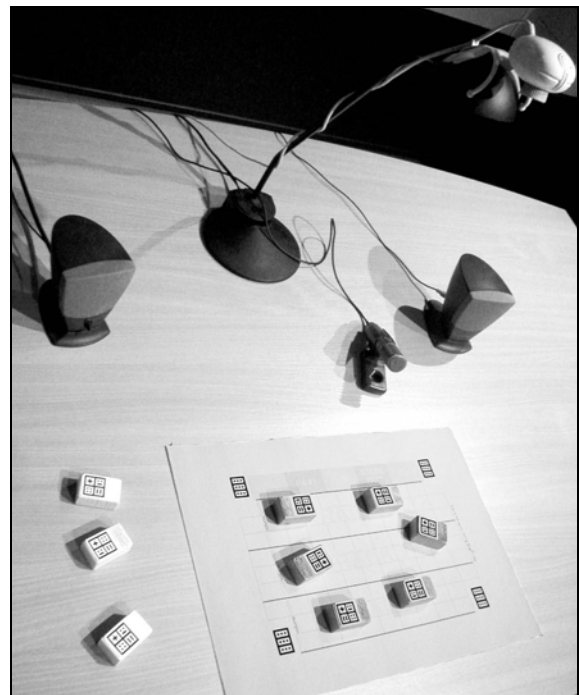


Figure 1: Entire set up of the system (Physical Sequencer).

2. DESCRIPTION OF THREE INSTRUMENTS

The general setup is the same for the three instruments: the interface is composed of an interactive area and a number of interactive blocks (Figure 1). The interactive area is a plane surface (e.g. a table top) observed by a web-cam connected to a consumer-

grade personal computer equipped with a sound card. The web cam is positioned on a table lamp (turned off and used as a cheap stand) pointing at the surface. The computer speakers need also to be in proximity of the interface to deliver the sound to the user. The computer itself and its monitor, however, do not need to be in view. In fact the user controls the system just by grasping and moving the blocks on the interactive surface. Feedback is provided both from the audio and from the physical arrangement of the blocks (which effectively act as both input and output devices). The interface area is covered with a sheet of paper which contains four special markers used to calibrate the system, and a number of visual cues for the user. These cues depend on the specific applications as described in the following subsections. In the current implementations the sheet of paper is an A4, but the entire system can be scaled to any size so long as the ratio between the interactive area size and the block size is constant.

1.2.1. Augmented Musical Stave

In the augmented stave, physical representations of musical notes can be placed on a stave drawn on an A4 sheet of paper to compose simple melodies or to teach the score notation to children. The interactive objects are rectangular blocks, about 8 by 3.2 by 1.7 centimetres in size. Each block is labelled with a fiducial symbol hidden in the shape of a musical note, or displayed next to a pause symbol. The system is shown in Figure 2.



Figure 2: *The Augmented Musical Stave.*

As soon as the notes are placed on the stave, the corresponding sounds are played by the computer. The pitch depends on the vertical position of the object on the musical stave, as in standard musical score notation. The objects/notes are available in different types, each representing a different duration (semi-quavers, quavers, crotchets, minims, semi-breves, etc...). The horizontal position of the note determines the sequence in which the sounds are played. When the end of the stave is reached the program loops back to the beginning of the stave, taking into account any changes in notes and their positions. Two or more notes of the same duration can be also arranged in a chord. The interactive area can be seen as a bar in terms of musical notation, the length of which is determined by the notes and rests that are placed within it.

2.2. Tangible Drum Machine

Loop based drum machines have proven very popular especially in the field of dance music. They allow users to produce drum beats and to some extent edit them as they are playing. The output of such a device can be connected to an amplifier or an audio mixer to be used in live performance as well as recording. However, the interfaces involved with such machines are often quite difficult to use and to make slight changes often involves a confusing sequence of button pressing.

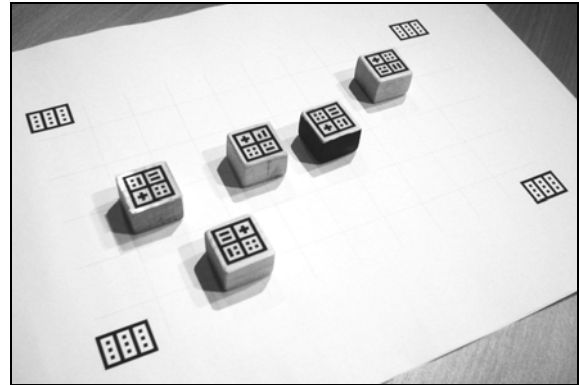


Figure 3: *The Tangible Drum Machine.*

The set-up for this application is similar to that of the augmented musical stave and it is shown in Figure 3. The augmented surface is again covered with an A4 sheet of paper. Rather than a musical stave, a grid is displayed to provide a visual cue for the user. This yields an interface where the user can build complex drum rhythms and naturally adjust them by moving the blocks.

This time the interactive objects (square based blocks about 2.8 by 2.8 by 2 centimetres in size) represent drum sounds in a loop. Their position on the horizontal axis determines the time they are played within the loop, while their position on the vertical axis determines different drum sounds. As with the stave, different types of objects are available. These different types of interactive object correspond to different sound volumes (to allow accents in the beat). This yields an interface where the user can build complex drum rhythms and naturally adjust them by moving the physical blocks. Compared to most drum machines our Tangible Drum Machine is very responsive to subtle adjustments in the timing of the samples. The resolution of the camera is sufficient to provide a fine time quantisation.

An additional feature that was noticed only after the implementation of the application is the ability to temporarily mute parts of the drum sequence by covering the fiducial symbols with a hand or similar object.

2.3. Physical Sequencer

In the physical sequencer the users can record live input connected to the sound card "on the blocks", and then physically arrange the

blocks on the interface to create a sequence which is continuously played in a loop. It is also possible to apply digital audio effects to the samples such as reverberation and change of playback speed.

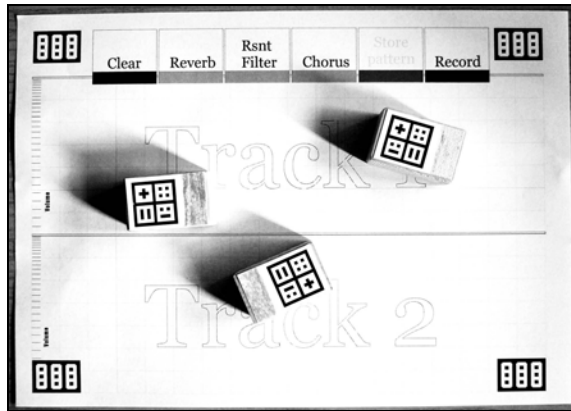


Figure 4: *The Physical Sequencer, overview.*

The interactive objects are wooden rectangular blocks of size about 2.5 by 4 by 2 centimetres. The blocks are marked with different colours, each colour corresponds to a different sound. The interactive area is divided in two main sections: the *tracks* area and the *control* area. Two tracks occupy the most of the interface. If a block is placed in one of the tracks its vertical position with respect to the track determines the volume, while (similarly to the drum machine) its horizontal position determines the trigger instant. The tracks have been replicated two times to allow different samples at the same time and at the same volume level. Multiple blocks of the same type can be used to repeat the same sample multiple times in a cycle. This feature can be used to build rhythmic patterns or to play with phase differences with longer samples. The number of tracks has been limited to two because of the limited size of the interface. It has been observed that even increasing the number to three becomes impractical, as the tracks become too narrow to allow an acceptable level of control over the volume of the sounds.

The "active areas" above the tracks allow the users to perform actions such as recording a new sample onto a block, or applying an audio effect to the sample; this is done by placing the block of interest on the relative area. There are five active areas. One is used for recording new samples, three are used to apply audio effects to individual samples and the final active area is used to remove any effects that have previously been applied to a sample. For the recording function, the current audio input is sampled until the block is removed from the active area. The recording does not start instantly, but only at the beginning of the next cycle. To reduce synchronization problems derived from this lag, any silence at the beginning of the sampled sound is automatically removed. An audio effect can be applied to a sample by placing a block in one of the three active designated for audio effects. The result is that the corresponding sound of that block (and all other blocks with the same label) is manipulated in real-time based on an audio processing algorithm. It is possible to apply more than one effect to the same sample by consecutively placing a block in more than

one audio effect area. The audio effects consist of a simple reverb implementation, a resonance filter and a chorus effect. The parameters of these effects are fixed for simplicity. The orientation of each block determines its playback speed, allowing some tone variation.

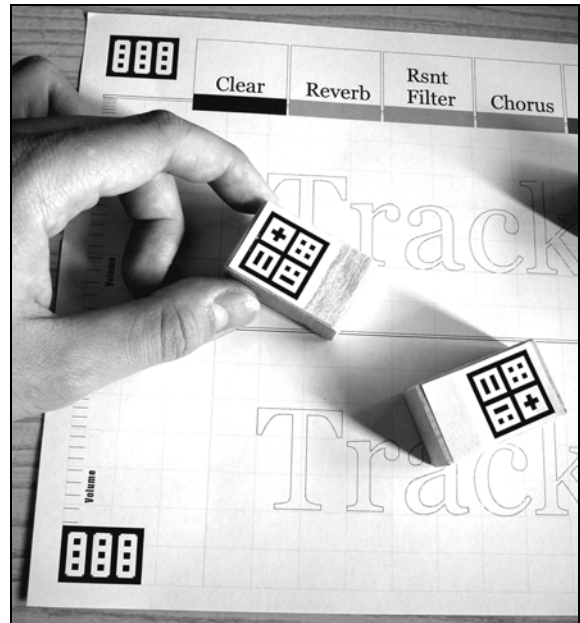


Figure 5: *The Physical Sequencer, detail.*

3. IMPLEMENTATION

The interactive objects are "seen" by the computer through the web-cam using the fiducial recognition system described in [9]. Each physical object is labelled with a special black and white symbol which is easily found by the computer. The system is fairly robust to illumination variations and shadows, even using a low cost web camera. The algorithm produces the coordinates and type (class) of each object seen by the camera. In the case of the instruments the block position is relative to the sheet of paper, which has four symbols printed on it. The fiducial recognition system has been implemented in the form of a toolkit for the development of tangible user interfaces, as reported in [4]. The vision algorithms are wrapped as a separate software layer.

The information produced by the fiducial recognition system is then used as input for the sound synthesis. All of the three configurations have been implemented with sampling synthesis using the Synthesis ToolKit (STK) [10], an open source set of classes for audio signal processing and algorithmic synthesis. One of the STK instrument classes – the *drummer* class – served as a starting point for the development of the sound engine of each of our configurations. A controller class is used to interface the fiducial recognition system with the sound synthesis class in a multi-thread environment. The sound engine was developed to

handle the polyphonic playback of musical samples. In the case of the Augmented Musical Stave and Tangible Drum Machine the samples are pre-set in the instrument and the engine performs only a playback function. For the Physical Sequencer, sounds are also recorded from the sound card input and stored on the disk. The same engine handles the application of audio effects to the individual samples. The effects used are also implemented using STK classes. The reverb effect – based on the 'JCREv' class – uses a series of three all-pass filters, followed by four parallel comb filters and two decorrelation delay lines in parallel at the output. The resonance filter was implemented using the 'BiQuad' class, which creates a two-pole, two-zero digital filter. This class is used to create a filter with a resonance in the frequency response while maintaining a constant filter gain. The STK class 'chorus' is used to implement the chorus effect.

This technique has allowed us to obtain the desired result with the time and hardware resources available. As reported in section 5, we are interested in the possibility of applying this user interface to other synthesis methods.

4. DISCUSSION

The approach presented can be compared with existing software based on GUIs. General advantages of TUIs have been discussed in [1, 2, 4, 11], and their benefits to musical applications are observed in [5, 6]. For example, using tangible interfaces as input allows "space multiplexing" rather than "time multiplexing" (as in GUIs). Different functions and devices are independently and simultaneously accessible. Hence, the interface can be used with both hands, and by several users at the same time. Furthermore, the simple mapping that we have chosen for our interface allows anyone to play and enjoy the experience of producing sounds, without requiring background computer knowledge.

4.1. Informal User Evaluation

A prototype of each instrument has been informally tested by a group of people with varied musical backgrounds; from a music academic to someone with little or no experience in music composition. Each enjoyed interacting with the instruments and were able to make interesting and varied compositions. We noticed that everyone who used the sequencer employed originality and imagination when choosing the source of the sounds they wanted to record, and seemed to be gripped by the mystery and exploration involved with producing different musical sequences with the samples. It was pointed out, however, that it is sometimes hard to tell what point of the loop was being played at any particular time. This could be resolved by some extra visual or audio feedback.

Our drum machine and sequencer provide a very fine quantization, encouraging the production of rhythms that are unusual and interesting. Some of the test users tried to compose audio loops arranging the blocks in geometric shapes (e.g. to "see what a triangle sounds like").

Users also noticed a number of advantages coming from not using a GUI and therefore not needing to stare at a computer monitor. Firstly there is simply the aesthetic appeal of using such a

system. People who were used to producing computer music enjoyed the fresh approach of controlling very precise information by manipulating simple blocks on a surface. Users were invited to experiment by closing their eyes and relying only on their senses of touch and hearing. They were still able to manipulate the sounds, and in some cases preferred this "blind" interaction. The three instruments have also been demonstrated to visually impaired users who enjoyed the interaction, despite the fact that the interface had not yet been optimized for this purpose.

4.2. Applications

The instruments can be used in different contexts, ranging from composition and performance to play and education. This is in virtue of their low cost, flexibility and scalability. The image processing approach allow the system to be implemented in different sizes, the only constraint arising from the camera resolution, which determines the ratio between the interactive area size and the object size. With a consumer grade web-cam (resolution 640x480 pixels) this ratio is about 100:1. In practical terms, a large scale interface can be used in a public space, as an installation or in an educational environment, a small scale one is portable and more suitable for a personal use.

Interest for musical interfaces for children has been expressed in [12]. Children from a very early age, before even learning to pick up a pen, would be able to create music by placing the objects on the stave, building a connection between the somewhat unnatural way in which music is represented in a musical score and the sounds that it represents. It has been suggested that the system could be a teaching aid for the Suzuki method. The instrument has been played and informally evaluated by four music educators and others involved in music therapy. Their feedback has lent weight to our belief that the interface will support student learning, and has suggested a number of lines of future development.

5. FUTURE DEVELOPMENTS

We are currently arranging formal subjective testing on the interfaces. To simplify the analysis of the results we plan to ask the testers to perform simple tasks using our interface and a similar GUI based application.

Further work will focus on enhancing the usability of the instruments. It has been noticed that the forced timing given by the automatic sweep without visual feedback confuses some users, especially those who are not familiar with music technology software. We also aim at optimizing the interface for visually impaired users moving from visual to tactile cues to distinguish different blocks. Elements on the interactive surface (such as active areas) should also be marked with tactile features.

As mentioned in section 3, we are interested in exploring the potentialities of the interface from the applications point of view (including use in music therapy). A particular area of interest would be to use the interface for a real time instrument, using different "active areas" and different blocks to control its parameters using appropriate mapping. The interface would then be generalized as a controller for musical applications. Such

development could be facilitated by linking the interface to a musical programming language such as PD or Max/MSP.

Other areas of investigation will cover the use of the interface as a communication medium in public spaces. For example musicians could set up the sequencer interface with audio samples. In an exhibition space then the audience will arrange blocks on the interface and listen to the piece. In this way the composition process is shared between the artist and the listener. The artist describes the content, but without completely fixing its form. The listener is thus able to engage actively with the artwork.

6. ACKNOWLEDGEMENTS

The authors would like to thank Dr A. Hunt at the University of York for the encouragement, help and refreshing discussions.

7. REFERENCES

- [1] Fitzmaurice, G., Ishii, H., Buxton, W., "Bricks: Laying the Foundations for Graspable User Interfaces", in *Proc. of the Conference on Human Factors in Comp. Sys. (CHI'95)*.
- [2] Ullmer, B., Ishii, I., "Emerging frameworks for tangible user interfaces", *IBM Systems Journal* 39(3&4): 915 2000.
- [3] Ishii, H., Ullmer, B., "Tangible bits: towards seamless interfaces between people, bits and atoms", in *CHI'97*, pages 234 -- 241, 1997.
- [4] Costanza, E., Shelley, S. B., Robinson, J., "d-touch: a Consumer-Grade Tangible Interface Module and Musical Applications" accepted for publication in *Proc. Designing for Society HCI2003*, Bath, UK, 8-12 September 2003
- [5] Patten, J., Recht, B., Ishii, H., "Audiopad: A Tag-based Interface for Musical Performance", in *Proceedings of Conference on New Interface for Musical Expression (NIME '02)*, Dublin, Ireland, May 24 - 26, 2002.
- [6] Newton-Dunn, H., Nakano, H., & Gibson, J., "Block Jam", in *Emergent Technologies of SIGGRAPH 2002*, p.67.
- [7] Poupyrev, I., "Augmented Groove: Collaborative Jamming in Augmented Reality", in *SIGGRAPH 2000 Conference Abstracts and Applications*, ACM Press, NY, p. 77
- [8] Hsiao, K., Paradiso, J., "A New Continuous Multimodal Musical Controller Using Wireless Magnetic Tags", in *Proc. of the 1999 International Comp. Music Conference*.
- [9] Costanza, E., Robinson, J., "A Region Adjacency Tree Approach to the Detection and Design of Fiducials", in *Proc. Vision, Video and Graphics*, 2003, Bath, UK, July 2003.
- [10] Cook, P. Scavone, G. P., "The Synthesis ToolKit (STK)", in *Proc. of the 1999 International Computer Music Conference*, Beijing, International Computer Music Association.
- [11] Patten, J., Ishii, H., "A Comparison of Spatial Organization Strategies in Graphical and Tangible User Interfaces", in *Proc. of Designing Augmented Reality Environments (DARE '00)*.
- [12] "DSP for children",
<http://www.notam02.no/~joranru/DSPforChildren.html>