MobiToss: A novel gesture based interface for creating and sharing mobile multimedia art on large public displays

Jürgen Scheible
Media Lab
University of Art and Design Helsinki
Hämeentie 135C, FI-00560 Helsinki
+358 40 550 9268
jscheib@uiah.fi

Timo Ojala
MediaTeam Oulu
University of Oulu
FI-90014 University of Oulu, Finland
+358 40 567 6646
timo.ojala@ee.oulu.fi

Paul Coulton
Informatics Group, Infolab21
Lancaster University
Lancaster
p.coulton@lancaster.ac.uk

ABSTRACT
This paper presents MobiToss, a novel application for creating and sharing mobile multimedia art with an off-the-shelf mobile phone equipped with built-in accelerometer sensors allowing gesture control. The user first takes a photo or captures a video with the phone and then using a 'throwing' gesture transfers the clip onto a large public display for instant viewing and manipulation by tilting the phone in different directions. The system augments the user-created clip with other items such as music or brand labels and the encoded clip is automatically sent back to the phone as a personal artefact of the event. The clip is also uploaded to a dedicated community website for sharing the created multimedia art with others. MobiToss could be deployed e.g. in clubs, pubs and concerts as a participatory VJ-tool. In addition to the design the paper presents the results of a preliminary user evaluation, which highlight the novel art experience provided by MobiToss.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation]:** User Interfaces – *evaluation and methodology, input devices and strategies, interaction styles.*

General Terms
Design, Experimentation, Human Factors.

Keywords
Gesture control, experimental evaluation, interactive content, public display.

1. INTRODUCTION
MobiToss combines standard interaction devices (mobile phone, large public display, web), interaction styles (mobile phone UI, gesture control) and multimedia processing techniques into a novel system for creating and sharing mobile multimedia art.

Of particular interest is the gesture control with a mobile phone equipped with built-in accelerometer sensors. Having shot a video with the phone the user can 'throw' the clip onto a large public display for instant viewing. Then the user can manipulate the video on the public display by tilting the phone in different directions.

Kela et al. [6] indicated in their study on accelerometer-based gesture control for a design environment that different people usually prefer different gestures for the same task, and hence it should be possible to personalize the gestures. By using the throwing gesture, MobiToss relies on a commonly understood gesture 'throwing something to somewhere'. However, the challenge is to design the system in such a way that different forms of throwing can be used reliably and intuitively.

Many people are used to throwing physical objects, but throwing digital media is a new concept. Bertelsen et al. [1] report a video prototype in which digital documents are positioned on walls and floors by throwing a ball at the wall or bouncing it onto the floor. They argue that the interaction with the ball serves to establish new relationships to the digital materials manipulated. MobiToss provides a starting point for exploring the throwing gesture in the context of digital media and collaborative screen based entertainment using a commercial mobile phone.

The Wii game console [8] introduced a new way of interacting with games on a large screen. While the Wii controller is just a gesture control device, a modern mobile phone has many additional features. It allows creating content as well as being wirelessly connected, and contains a built-in accelerometer sensor for gesture control (e.g. Nokia N95 and Nokia 5500). MobiToss combines these features into a single application.

Vajk et al. [10] presented a system that connects the mobile phone to large public game screens over Bluetooth and turns the phone into a game controller in a multiplayer game. They highlight that this type of control can be both intuitive and fun. Similarly, SQEAK [7] employed mobile phones with attached miniature accelerometers for gesture based group interaction with public displays.

While exploring the use of a mobile device as an artistic tool in public spaces, Clay et al. [2] concluded in their mobile music work 'China Gates' that using wearable computing technology

MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.

within global ubiquitous networks as an art tool allows interacting with society as part of a collective.

This paper is organized as follows. Section 2 describes the MobiToss system. Section 3 presents the initial experimental evaluation of the system in the true environment of use and discusses the important findings and future work. Section 4 concludes the paper.

2. THE MOBITOSS SYSTEM

The user first captures a photo or video with a personal mobile phone and then 'throws' it onto the large public display by making a throwing gesture with the mobile phone, as illustrated in Figure 1. Upon a 'recognised' throw a matching sound is provided by the phone as feedback for success and the content is instantly visible on the screen.

Figure 1. Throwing a video to the large screen.

The user can manipulate the appearance of the video or photo on the screen by moving the phone up, down, left or right, as shown in Figure 2. The movements correspond to different video effects and graphical filters such as matrix, kaleidoscope, spiral distortion and color saturation. For example the matrix effect multiplies the video over a number of rows and columns and by tilting the phone forward or backward, this number can be increased or decreased, which leads to a zoom effect. Using the spiral distortion effect the user can make the entire content rotate around the centre of the screen, or via the color saturation effect all colors of the video can be saturated by the same degree etc. The video effects provide a means for creating artistic outcomes but could also be used as built-in censorship mechanism for inappropriate content.

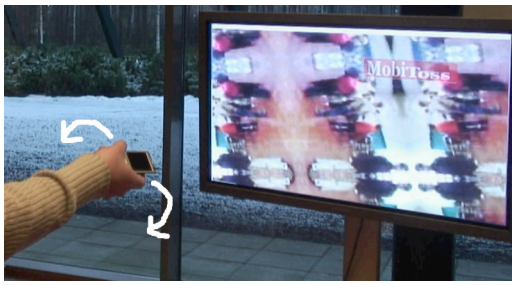


Figure 2. Gesture control to manipulate video.

The produced artefact, viewable by everyone, is recorded into a new 30-second video clip and encoded with a commercial pop song, a logo and a web address as advertisement. The encoded mash-up clip is then automatically uploaded to the users' phone as a personal reward and memory of the created artefact (Figure 3). The mash-up clip is also uploaded to a dedicated community website, to collect the clips and to make them available to other users and to share them with friends. The website can be configured for creating event and location based collections.

There are several reasons why the mobile phone is a highly suitable user device for MobiToss. First, as the personal trusted device it stimulates use and removes need for any proprietary application specific devices. Second, there are already several phones with built-in accelerometer sensors on the market and they are predicted to become increasingly widespread. Third, they provide the functionality to create content by the users and allow anonymous, wireless participation in a joint social public group interaction. Fourth, the mobile phone provides a reliable return channel for delivering confidential user specific content back to the user, such as the music video including logos and web links as advertisement.

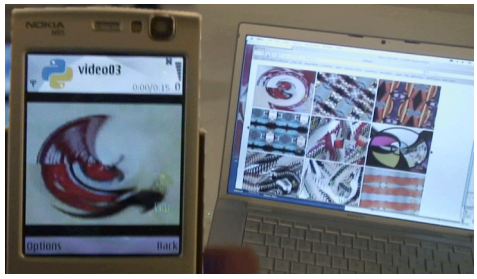


Figure 3. Mash-up clip on phone and community website.

The MobiToss system is realized with a client-server architecture, which comprises of four components: 1. A Symbian client application running on a mobile phone with built-in accelerometer. 2. A server running on a PC for creating the clips. 3. A large public display showing the video. 4. A dedicated community website for collecting the clips.

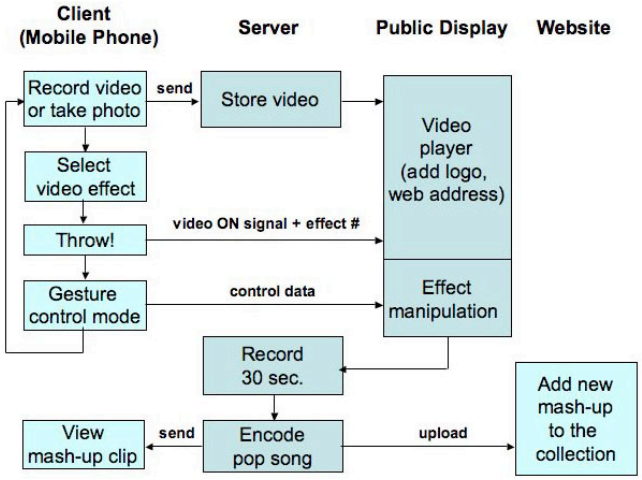
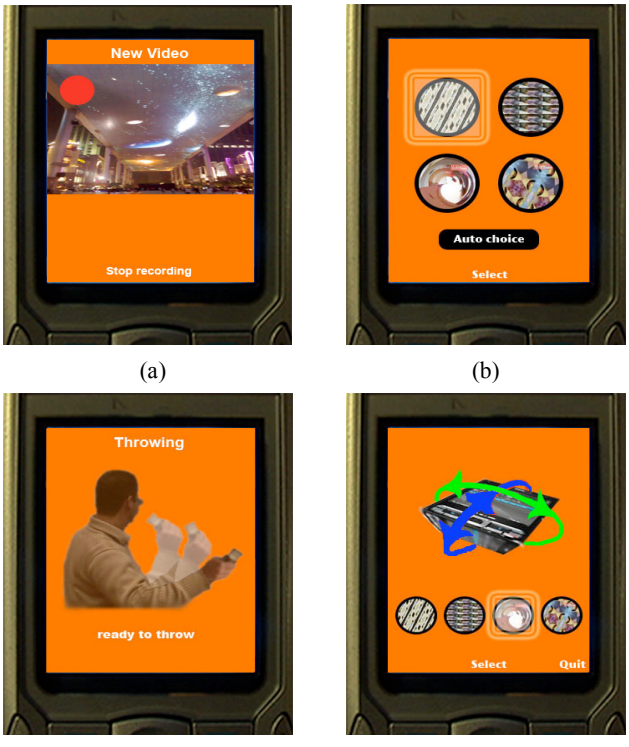


Figure 4. State diagram of the MobiToss system.

The client is implemented in Python for S60 [9] that allows the reading of the built-in accelerometer sensor data as well as shooting of photos and videos by a single application. The client also executes a range of tasks automatically in the background

such as uploading/downloading of content or controlling the video manipulation on the server. The client communicates with the server using the HTTP protocol and sockets over WLAN (IEEE 802.11) connection. Bluetooth connectivity is not used since it would restrict participation to a limited number of users (8 max) within a limited spatial range of about 10 meter radius from the service point. Using WLAN provides sufficient data rate and allows participation from much greater distances, for example throughout a concert held in a large arena.

Figure 4 shows the state diagram of the system. First, the user chooses from a menu to either take a photo or to record a video. The selection invokes automatically the camera's viewfinder. The recording period is set by pressing the start and end buttons (Figure 5(a)). The phone then sends the video to the server where it is stored and handed to the video player of the public display. Next, the user is provided with a menu to select a video effect (Figure 5(b)) for manipulating the video on the public display. Each effect comprises of a set of parameters for later remote manipulation of the video's appearance. The selection of the video effect at this point serves the purpose of giving the user control over the initial visual appearance of his or her video before it can be seen on the public screen.



(a) (b)

(c) (d)

Figure 5. Screenshots of the client UI: (a) the video recording has started; (b) video effect to be selected; (c) the application is ready for throwing the video; (d) gesture control mode is active.

Once the sending of the video is complete (measured in seconds), the phone notifies the user that (s)he can now throw the video by playing a sound and by displaying a 'Ready to throw' message (Figure 5(c)).

When a throwing motion of the phone (Figure 1) results in accelerometer readings exceeding certain thresholds, the client sends an HTTP request to the server to make the video visible ('video ON signal'). The client switches immediately into gesture control mode (Figure 5(d)) and uses socket communication with the server to transmit control signals for real-time manipulation of the video. The control bytes are generated from the accelerometer readings reflecting the phone movements such as tilting left/right or forward/backward (Figure 2). This is the time when the user can create his or her visual art piece on the large screen. The creative experience is enhanced by the possibility to use the phone's navigation keys to switch in real-time between different video effects (Figure 5(d)). When the system enters the gesture control mode the server starts recording a 30-second a clip of the video as it appears on the public screen. Once the recording is complete, the server encodes a pop song into the clip. The encoded clip is automatically sent back to the phone and uploaded to the community website.

The server implemented in Max/MSP Jitter [5] handles all video effects and adds graphic elements such as brand icons and web links to the video. It also records the 30-second video. Additional components include the ffmpeg software [3] for adding the music into the clip, and some simple Python scripts for connectivity and server management.

The large display with sufficient presentation capacity serves as the main public user interface for the group, while the mobile phone provides a private GUI and return channel to the user. The community website (Figure 3) is there to collect the final clips. Anytime a new clip is generated it is automatically uploaded to the website. This provides a channel to make the clips available to other users, and to share them with friends. The MobiToss system can be configured so that it can create event or location based collections.

3. USER EVALUATION

The first evaluation of MobiToss with real users was carried out at a social event of an international conference. The goal was to obtain first impressions on how the system is perceived by the users and to identify the strengths and weaknesses as input for the next design cycle. As shown in Figure 2, the public display was located 2-5 meters in front of the users. Each user was given a Nokia N95 phone containing the client application, and shown how to start the client. Approximately 25 people, most of them researchers on mobile multimedia, tested the system and feedback was collected from 11 persons via video interviews.

In the following, we present the issues that caught our's and test user's attention. As there is insufficient space in this paper to analyze these issues in detail, significant factors are presented in the form of the answers and comments given by test users in response to a series of questions by the researchers. *Test user's comments are printed in italic.*

A clear general observation was that people enjoyed the idea of 'throwing' and playing with the system. Question: How was your experience using MobiToss? *"The throwing is the best, I like it". "It is very reactive, I didn't expect this reaction, it works well."; "It is pretty nice tool for a social event like this because you can capture the event live and you can have some fun. I like that it meshes the technological stuff with the artistic effects"; "It's fun to recognize other people in the video."*

UI issues. We observed that different people do their throwing movement in many different ways, confirming Kela's finding [6]. The current implementation did not successfully detect all users' throwing gesture. User feedback brought up ideas for developing the UI further: *"Have on the phone screen an animated preview of the possible effects to see what they do"; "Reduce the number of video effects, probably have only 2-3 very clear ones to explore the system, and maybe you are more immersed.";* This indicates a need for careful UI redesign, making the relationship more clear between effect parameters and the interaction.

Motivation for participation. MobiToss succeeds in providing a tool for artistic expression. *"Having the feeling I can produce my own pieces of art"; "If I could impress other people with doing the thing, than I would be ready to pay e.g. 50 cent"; "I would put the faces of my friends and my pet there."; "To get the music piece together with the ready art piece sent your phone is nice*". It seems people need to get something out of it when using MobiToss, and doing something artistic appears to fill in on this.

Immersion and perception issues. Whereas using MobiToss was a fun experience to many, it became obvious that the perception of what is going on and what interaction causes what, couldn't be fully grasped. *"I didn't really feel immersed. For me it would mean that I am attracted to the screen while I am interacting, so that I have the feeling that even I can contribute to the evolving piece of Art"; "It didn't have as much meaning as it would have if you would be able to identify things more clearly in the video"; "One has to try it out few times to figure it out.*" This indicates that the system needs still a careful redesign of the video effects and its manipulation parameters.

Commercial aspects. We were interested to know if users would be ready to pay money for downloading the client application. *"It depends, if I download it once and use as much as I want, then it is ok for me to pay. But if you would have to pay every time you go to a party, or if its time based payment, then probably not."; "It would be better if you can customize the filters and effects and give the people options to buy them separately. It would depend on the quality and how artistic they are on how much they cost."* It appears that some people would be ready to pay a small amount and exploring various payment models might be feasible.

Collaboration aspects. The participants' feedback speaks in favour of having more than one person to throw their video and interact with the screen at the same time: *"Maybe you could have a competition with a split screen and you can compare the two players outcome in real-time."; "When several people throw their video - depending on the starting time of throwing - they can collide and depending on the crash a new thing is created.";* To bring in elements of competition and group play is certainly enriching for MobiToss and should therefore be integrated in the future.

Advertisement. MobiToss can encode branding content as well as music into the user-generated mobile multimedia art clip that resides in peoples phones but also on the public sharing website. This offers great opportunities for placing advertisements in a new way.

4. CONCLUSIONS

This work shows that capturing and throwing mobile content onto a large screen and manipulating it with gesture control into an art piece was perceived as a fun activity. The concepts of "me the artist" and "my art piece creation" are very relevant to the users of MobiToss. Having these concepts combined with the social aspect as well as the fact that the audience can join easily, makes MobiToss a potential new form of input to mobile community websites, e.g. as a participatory VJ tool or a public performance system. Community websites such as YouTube, Flickr or SeeMeTV have become important parts of the Internet, combining user-generated content with advertising. The so-called GENERATION C phenomenon [4] is emerging. It refers to the avalanche of user-generated 'content' in the web where terra-bytes of new texts, images, audio and video are accumulated at an ever increasing rate. The underlying driver seems to be that we are all artists, but until now we neither had the personal drive nor the means to pursue this goal. Therefore, there is a need for better tools to create, to produce, and to participate. And systems such as MobiToss could serve as new types of collaborative tools for creating novel forms of input and participation to community websites, possibly coupled with commercial content.

Although MobiToss worked reasonably well in terms of engaging users and their creativity, it requires further development in order to give the users the intended full-scale experience, e.g. by applying a more balanced set of video effects, adding group interaction and a more intuitive UI. Further, the system can also be expanded in different ways, e.g. into a collaborative tool in meeting or class rooms, where everyone can just throw their content to the large screen and control it from their own seat, or into a remote browser for photo, video or other content.

5. REFERENCES

[1] Bertelsen, O.W., Petersen, M.G. and Pold S. (Eds.), 2004, Aesthetic Approaches to Human-Computer Interaction Proceedings of the NordiCHI 2004 Workshop.

[2] Clay, A., Frey, T., Gutknecht, J. and Majoe, D. 2006, Motion Tracking Music with Mobile-Media Technology in Two Case Studies. In Proceedings of the International Symposium on Pervasive Computing and Applications 2006, 358-363.

[3] ffmpeg: http://ffmpeg.mplayerhq.hu/.

[4] Generation C. http://en.wikipedia.org/wiki/Generation_C.

[5] Jitter: http://www.cycling74.com/products/jitter.

[6] Kela, J., Korpipää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L. and Marca, S.D. 2006 Accelerometer-based gesture control for a design environment. Personal and Ubiquitous Computing Journal, Springer, 285-299.

[7] Majoe, D., Schubiger, S., Clay, A. and Arisona, S.M. 2007, SQEAK: A Mobile Multi Platform Phone and Networks Gesture Sensor. In Proceedings of the Pervasive Computing and Applications conference ICPCA 2007, 699-704.

[8] Nintendo wii remote. http://wii.nintendo.com.

[9] Scheible J., Tuulos V. 2007. Mobile Python -- Rapid prototyping of applications on the mobile platform, John Wiley & Sons.

[10] Vajk, T., Coulton, P., Bamford, W. and Edwards, R., 2008, Using a Mobile Phone as a "Wii-like" Controller for Playing Games on a Large Public Display, International Journal of Computer Games Technology, Volume 2008.