

The Tangible Video Editor: Collaborative Video Editing with Active Tokens

Jamie Zigelbaum¹

Michael S. Horn²

Orit Shaer²

Robert J.K. Jacob²

¹MIT Media Lab
Tangible Media Group
20 Ames St. Cambridge, Mass. 02139
zig@media.mit.edu

²Tufts University
Department of Computer Science
161 College Ave. Medford, Mass. 02155
{mhorn01, oshaer, jacob}@cs.tufts.edu

ABSTRACT

In this paper we introduce the *Tangible Video Editor* (TVE), a multi-user, tangible interface for sequencing digital video. We present a new approach to tabletop interaction by using multiple handheld computers embedded in plastic tokens. Drawing from the rich physical experience of traditional film editing techniques we designed the TVE to engage multiple users in a collaborative process and encourage the exploration of narrative ideas. We used *active tokens* to provide a malleable interface, enabling users to organize the interface components in unspecified ways. Our implementation improves upon common projection-based tabletop interfaces in a number of ways including a design for use beyond dedicated two dimensional spaces and a naturally scaling screen resolution.

Author Keywords

Tangible user interface, digital video editing, active tokens, interface design, CSCW, distributed cognition, tabletop interaction, physical interaction.

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

INTRODUCTION

Traditional film editing systems such as Moviola and Steenbeck provide editors with a rich sensory environment that allows them to utilize enactive (muscle) memory and haptic (force) feedback. The task of cutting and splicing film into sequences for playback involves the use of physical tools such as cutting arms, and taping stations. The

affordances of these tools help convey both their purpose and their means of use. In contrast, current state-of-the-art, NLE (non-linear editor) software such as Final Cut Pro, Premiere, and AVID provide filmmakers with little of the physical richness employed by their predecessors. However, they provide editors with powerful features such as the ability to reuse shots without making a new print from a negative, to undo actions as quickly as they were done, and to initiate new projects with the click of a button.



Figure 1. Working with the Tangible Video Editor.

The Tangible Video Editor (TVE) project is, in part, an attempt to combine the benefits of traditional, physical film editing with the advantages of digital, non-linear film editing. The implementation of the TVE presented here is a tool for basic video editing and storytelling, designed to engage multiple users in a collaborative process, and encourage the exploration of narrative ideas. We developed an *active token* approach to tabletop interaction to support this goal and created a malleable user interface that improves upon some of the limitations we found in similar projection-based interfaces.

RELATED WORK

There are many video editing interfaces in the HCI literature, and a few that use physical interactions for related tasks. An early example of a physical interface that could be used for editing video is Video Mosaic by Mackay and Pagani [7]. Essentially an enhanced storyboarding tool, Video Mosaic addresses the issue of representing time spatially by combining the benefits of real-world interaction (in this case with paper) and the power of digital manipulation. A more recent example is the mediaBlocks [11] system, which provides an interface for capturing, editing, and displaying multi-media content using passive wooden blocks as tokens. Although mediaBlocks is not a video editing system, the media browser and media sequencer functions are particularly relevant to our work, they allow users to view and sequence digital video by arranging the blocks within a physical device.

Tangible Video Browser [10] is another interface that uses physical tokens to represent video clips. When the tokens are placed on the browser interface, they become objects that can be manipulated to navigate through the clip itself. TellTale [1] and EnhancedMovie [6] are two other relevant interfaces. TellTale, a toy caterpillar whose body is made up of segments that can be arranged in any order, is an interface for creating sequences of audio clips. EnhancedMovie features an augmented desk, which allows users to make a movie by editing a sequence of pictures using hand gestures. Hand gestures (such as closing all fingers above an image) allow users to grab, release, select, and browse clips displayed as images on the table. Finally, Moving Pictures [13] is a tangible interface aimed at allowing young users to collaboratively create, manipulate and share video content. The system consists of a video station, a screen, an interactive table, and a set of tokens. To create a movie, users can move between three modes of operation: Shooting, VideoJockey and Storyboard. Each of these modes supports collaborative physical interaction guided by a GUI.

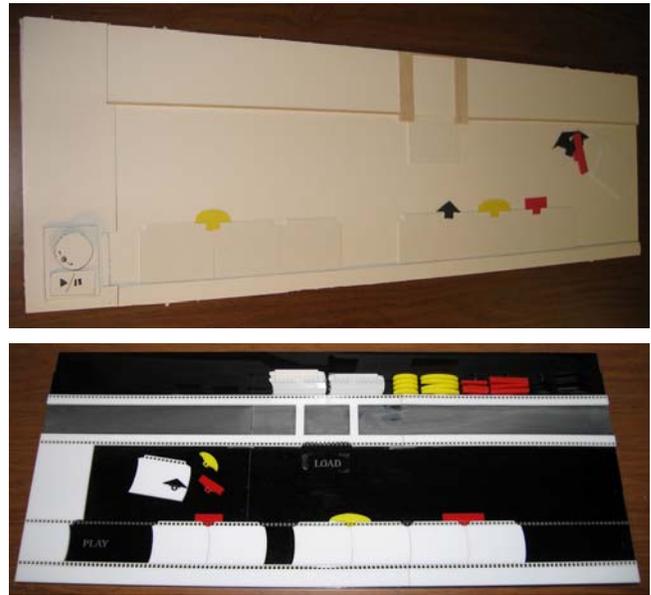


Figure 2. Foam core and plastic prototypes.

DESIGN PROCESS

The current version of the TVE is the result of an iterative design process where we constructed and tested multiple prototypes with user groups. Our first studies took the form of low-fidelity prototypes made from paper, foam core, and laser-cut acrylic (Figure 2). We used principles of tangible interaction in our design as they are widely cited to aid in collaboration [1, 5, 9]. After testing these prototypes along with another implementation using standard projector and computer vision techniques, we hypothesized that a reduction of global constraints would foster a freedom of interaction that would aid collaboration. In short we wanted to provide an interface that users could organize and manipulate in their own ways [4]. Projector-based systems have limited resolution and confine users to a predefined workspace and set of interactions, so we developed a new interaction technique based on active tokens.

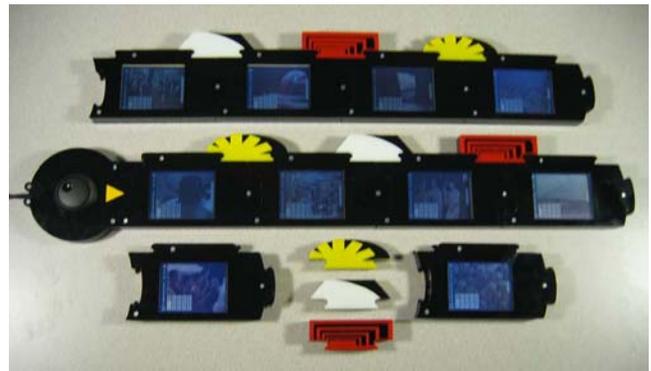


Figure 3. The components of the TVE.

IMPLEMENTATION

The current version of the TVE features a *play-controller*, 10 *clip-holders*, and 15 *transition connectors* (Figure 3).

These components operate in conjunction with a desktop PC and a monitor for viewing video content. Each clip-holder consists of a Compaq iPaq Pocket PC mounted inside a plastic case. By using multiple small screens, the TVE interfaces' resolution increases with the addition of each clip.

Physical Syntax

Designers of tangible interfaces often use form and mechanical constraints [12, 8] to express digital syntax. We designed the TVE around a jigsaw puzzle metaphor to offer both a cultural hint and a physical constraint that suggests that components can be connected in a sequence [2]. Furthermore, it is not necessary for users to learn special commands to perform operations such as removing a clip from a sequence [3].

Case Design

The cases for the clip-holders, transitions, and play-controller (Figure 4) were constructed from layers of 1/8 inch thick extruded acrylic sheets, cut with an industrial laser cutter. The iPaqs were removed from their original cases and mounted inside the clip-holders. A top layer of acrylic holds the iPaq in place and hides its internal components. Only the touch screen display is visible to the user. Copper connectors run along the outside edges of the cases where the clip-holders interlock with transitions and other clip-holders. The copper connectors are wired to the audio-out, audio-in, and application shortcut buttons of the iPaqs. When two clip-holders are joined, an electrical connection is formed between the audio out channel of the right clip-holder and the audio in channel of the left clip-holder. Likewise, when a transition is placed between two clip-holders, it closes a circuit to press one of the iPaq's application shortcut buttons. The software running on the iPaqs registers these events and inserts the corresponding transition data into the clip-holder data stream. The TVE supports three types of transitions: minimize, rotate, and fade.

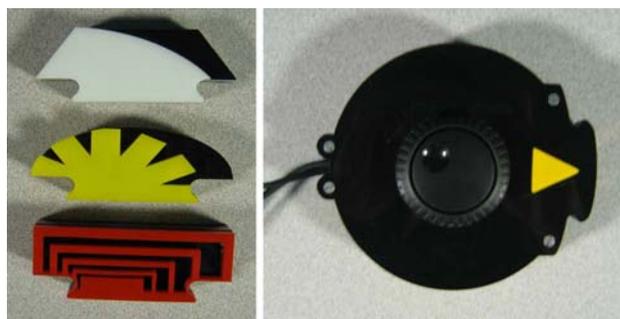


Figure 4. Counterclockwise from top: clip-holder, transitions (fade, rotate, minimize), and the play-controller.

Play-Controller and PC

The play-controller is a circular device that includes a yellow play button and a jog wheel. Users connect the play-controller to a clip-holder or the beginning of a sequence of clips holders that they want to view. Pressing the play button forwards the data stream to the desktop PC and triggers an application to play the clip sequence with transition effects on the desktop monitor. The application, written using Macromedia Flash, dynamically composes the movie from pre-loaded video clips.

Clip-to-Clip Communication

A clip-holder must be able to pass information to adjacent clip-holders about its current clip ID number and transition. This information flows along in a sequence from the rightmost clip-holder to the leftmost clip-holder, through the play-controller to the PC. Each clip-holder in the sequence receives a data stream from its neighbor to the right, appends information about its own clip and transition, and then passes the new data stream to its left. This data stream is transmitted in a continuous loop. We use the iPaq's audio channels (microphone and speaker jacks) to encode ASCII data in a frequency modulation format.

OBSERVATIONS

We observed 36 subjects working in pairs to evaluate whether the TVE provides benefits in supporting collaborative editing, users' engagement, and exploration of alternative narratives. Eighteen subjects interacted with the TVE and the other 18 subjects interacted with a typical graphical video editor. The subjects were 18-74 years old and were not paid. They came from a variety of backgrounds including college students, professionals, and retired persons.

The subjects' task was to work with a partner to create a short movie from pre-existing video clips. We provided 28 short video clips for participants to work with and three types of transitions to add between clips. Participants were randomly assigned to one of two editors: the TVE and Microsoft Movie Maker (MMM). MMM is a typical non-linear editing application for home video. Participants filled out both a pre-test and a post-test questionnaire about the task and the editing process. Each session was filmed for further analysis.

Results

Overall results were positive for the TVE. Subjects reported that they were more excited about creating their films and more enthusiastic about the editing process in the TVE group than the MMM group. They also reported that they had more fun.

During our observations we noticed very different behaviors between groups working on the two interfaces. As expected we found that teams working with the TVE shared the work more evenly. Our analysis of the session video recordings showed that subjects in these teams took more turns speaking than MMM subjects, i.e. one team member did not dominate the conversation in the TVE condition as frequently as in the MMM condition. Subjects working with MMM would most often designate one person to operate the keyboard and mouse for the duration of the task. TVE subjects, on the other hand, most often shared all of the work, including operation of the play button.

An unexpected behavioral difference between the groups became apparent within the first few minutes of each session. Subjects working with MMM often began the sessions in silence watching the available clips one after another where subjects working with the TVE often began the sessions discussing possible story lines together and organizing subsets of clips in their work space. Subjects using the TVE organized clip-holders in their workspaces in meaningful ways to add a layer of information, e.g. clips nearer to their bodies were often more important.

MMM users reported that they became more familiar with the content of every video clip than participants working with the TVE. This was confirmed by our observations—we noticed that TVE users would often view only half to three quarters of the available clips, and these they rarely watched more than once. MMM users would often watch the clips over and over again, and the teams usually watched every available clip. This was probably due to the fact that users of MMM only had to double click on a thumbnail of a clip to watch it where TVE users had to attach a clip holder to the play controller to view it. Although the TVE users did not view all of the clips they began discussing their narrative very early and explored multiple possibilities. One of our goals for the TVE was to encourage this type of exploration, so we were happy to see this during the evaluation.

Some of the users complained that the TVE did not offer enough functionality. Their main suggestion was to add the ability to cut and merge clips. Other common suggestions were to add audio editing and more of the functions found in GUI editors such as color correction, cropping, and speed control.

CONCLUSION AND FUTURE WORK

In this paper we have introduced the Tangible Video Editor (TVE), a tangible interface for editing sequences of digital

video clips. The TVE demonstrates a new implementation for tabletop interactions that uses active tokens. By embedding handheld computers in the interface components the TVE can be used on any surface under various lighting conditions and without the need for an augmented surface, computer vision system, or video projector.

Our observations of users working with the TVE during a collaborative editing task suggest that our departure from projection-based tabletop interactions was successful and warrants further investigation. We were particularly interested in the way users arranged the TVE components in their workspace and believe that this behavior suggests that the malleability of the TVE interface was beneficial. Our observations thus far have been subjective and although there is evidence supporting our use of active tokens, formal investigations would be valuable.

We have found this version of the TVE to function well as a simple device for video editing. In the future we plan on adding functionality to the system so that it can be used for more complicated editing tasks. Future refinements to our interface, along with the development of new interfaces for related tasks, will allow us to demonstrate more clearly the value of active token implementations for collaborative tabletop interactions.

REFERENCES

1. Ananny, M., Supporting children's collaborative authoring: practicing written literacy while composing oral texts. In *Proc. Computer Support for Collaborative Learning 2002*, Lawrence Erlbaum Assoc. (2002), 595–596.
2. Fishkin, K., Gujar, A., Harrison, B., Moran, T., and Want, R. Embodied user interfaces for really direct manipulation. In *Communications of the ACM*, 43, 9, (Sep. 2000), ACM Press (2000), 74–80.
3. Gutwin, C. and Greenberg, S., Design for individuals, design for groups: tradeoffs between power and workspace awareness, In *Proc. CSCW 1998*, ACM Press (1998), 207–216.
4. Hollan, J., Hutchins, E., and Kirsh, D., Distributed cognition: towards a new foundation for Human-Computer Interaction research, In *TOCHI*, 7, 2, (June 2000), 174–196.
5. Hornecker, E., Understanding the benefits of graspable interfaces for cooperative use. In *Proc. of Cooperative Systems Design 2002*, 71–87.
6. Ishii, Y., Nakanishi, Y., Koike, H., Oka, K., and Sato, Y., EnhancedMovie: movie editing on an augmented desk. In *Adjunctive Proc. of Ubicomp 2003*.
7. Mackay, W. and Pagani, D., Video Mosaic: laying out time in a physical space. In *Proc. of the Second ACM International Conference on Multimedia*, ACM Press (1994), 165–172.

8. Shaer, O., Leland, N., Calvillo-Gamez, E. H., and Jacob, R. J., The TAC paradigm: specifying tangible user interfaces. In *Personal and Ubiquitous Computing* 8, 5, (Sep. 2004), 359–369.
9. Suzuki, H. and Kato, H., AlgoBlock: A tangible programming language, a tool for collaborative learning. In *Proc. of 4th European Logo Conference* 1993, 297–303.
10. Tangible Video Browser.
<http://tangible.media.mit.edu/projects./tvb/>.
11. Ullmer, B., Ishii, H., and Glas, D., mediaBlocks: physical containers, transports, and controls for online media, In *Proc. SIGGRAPH* 1998, ACM Press (1998), 379–386.
12. Ullmer, B., Ishii, H., and Jacob, R. J., Token+constraint systems for tangible interaction with digital information, In *TOCHI* 2005, ACM Press (2005), 81–118.
13. Vaucelle, C., Africano, D., Davenport, G., Wiberg, M., and Fjellstrom, O., Moving pictures: Looking Out/Looking In, In *SIGGRAPH* 2005, ACM Press (2005).