DESIGNING ENVIRONMENTS FOR COLLABORATIVE VIDEO EDITING

Lucia Terrenghi[1], Torsten Fritsche[2], Andreas Butz[2]

Vodafone GROUP R&D
Chiemgaustrasse 116, 81549 München
lucia.terrenghi@vodafone.com

Media Informatics Group,
University of Munich, Amalienstr. 17, D-80333 Munich,
info@torstenfritsche.de, butz@ifi.lmu.de
www.medien.ifi.lmu.de

Keywords: Video, mobile, tabletop, design, evaluation

Abstract

Video capture has become very easy and happens spontaneously with current mobile phones, but the resulting clips are never edited, mostly because this would imply a different physical and social context. We present the design of an interactive tabletop application for collaborative video editing, which aims at preserving this spontaneous attitude and the spirit of sharing. It uses fluent and two-handed interactions, as well as physical interface objects. The prototype we built was evaluated in a focus group of teenagers, which confirmed many of our design decisions, but also raised new questions.

1 Introduction

Digital Video has become very accessible with the widespread use of camera-equipped mobile phones, and research has started to analyze the process of "videowork" [8], i.e., how people engage with video technology in general. Particularly teenage users tend to take a very lightweight approach to this. They often just capture short video clips with their mobile phone or other multi-function devices, and share these clips momentarily on the mobile screen or via Bluetooth, or later on websites, such as YouTube. Their focus clearly is on sharing, while editing does not seem to be a major requirement for this group of users.

Of course, there is a different possible perspective on this: Whenever a technology is designed for a novel interaction scenario, it might just not be designed appropriately to support the users' main goals, in this case shareability and enhancement of social relationships. The current – potentially bad – design then also affects the users' imagination of future ways of interaction and hence influences the apparent requirements. Current tools for video editing only provide a single-user experience on a desktop PC, in software environments, which support goal-driven interactions. This is in obvious contrast to the social motivation driving teenagers to engage with video technologies in the first place.

In this paper, we address the potential mismatch between teenagers' shareability goals for videowork and the available tools for video editing by designing a playful appliance for collaborative video editing on a digital tabletop. Teenagers' communication goals, indeed, makes them potentially a suitable target group for applications supporting collaborative creativity. Our main idea here was to enhance the kind of creativity that Gold [4] describes as "for the pleasure of sharing and giving". This is comparable to the pleasure of singing a song to a child, or preparing a meal for guests. In this sense, we are trying to provoke creativity by providing an opportunity for engagement and social communication.

Two main technological trends open up novel possibilities for the design of scenarios for collaborative creativity: On the one hand, the wide distribution of small, portable devices (e.g., camera-equipped mobile phones), capable of capturing, transferring, and retrieving different media types, have caused a democratization of content creation, communication, and consumption. In particular, these technologies provide entire user groups, such as teenagers, with novel possibilities for capturing and sharing their experiences, thus shaping their communication habits and relationships. On the other hand, large multi-touch displays, situated and embedded in furniture such as tabletops and whiteboards, allow the parallel interaction of multiple users (e.g., Microsoft Surface™ [11]). This paper reflects on the potential of bridging these two interaction contexts – small distributed devices and large interactive surfaces – in order to create opportunities and environments for collaborative creativity. By designing an application for collaborative videowork on a digital tabletop, we exemplify the challenges that the design of such environments needs to face in order to successfully support teenagers' practices and enhance their creativity and communication.

In a first step, we will consider how different device form factors afford different stages and shareability of videowork. In particular, we will discuss how large interactive surfaces can afford different levels of shareability, and consider how design can enhance those. We will then present our design goals and decisions, in consideration of related work. We will describe how the design was implemented in a prototype application and evaluated in a focus group with teenagers. The results will be discussed in light of their implications for the design of future, similar applications.

2 Small devices and videowork

Small devices allow the ad-hoc, spontaneous capture of digital video, and its transmission to other devices. Video sharing can simply happen by literally watching over other people's shoulder because of the device's form factor and its single-user affordances. While the screen size and the limited interactions afforded by these devices also limit the possibilities for video editing, digital video eases storage and distribution of the content. Sharing can hence also happen remotely, e.g., by uploading the clips to a website. In both sharing scenarios, no editing is normally performed. Creativity, in this sense, is only happening at the time of capture, and it is an individual arbitrary action. The following sections will explore the potential of affording different types of shareability and creativity by exploiting the form factor of other digital display categories, such as interactive tabletops.

3 Large interactive surfaces for collaborative video editing

Large surfaces, such as walls and tabletops, provide inherent social affordances by making the displayed information visible to multiple individuals. Furthermore, large displays also afford the concurrent visualization of a large amount of information, and thus provide the possibility to simultaneously see and compare multiple alternatives. The combination of these two aspects (i.e., multiple observers and multiple displayed alternatives) normally offers a rich environment for the collaborative creation, manipulation, and assessment of different alternatives in mundane creative processes.

The possibility of making such surfaces interactive and displaying digital content on them creates interesting and challenging opportunities for the design of social interactions across digital and physical media in space. Some main technologies, in particular, provide novel possibilities for the design of social user experiences:

- multi-touch, i.e., space-multiplexed input, allowing parallel interaction of multiple users, and/or two-handed interaction on interactive surfaces (e.g. [11]);
- object sensing, allowing the use of physical objects - and their manipulation vocabulary - for the manipulation of digital media on a surface (e.g. [5][7]),
- networked displays, allowing the design of interactions which unfold across multiple displays and devices, e.g. tabletops and mobile devices (e.g. [6]).

Making shared artefacts and spaces interactive by these technologies and controllable by all co-located users enables novel forms of shareability.
Based on the individuals' rights of parallel and co-located interaction with the information content, different levels of *shareability* can be distinguished, such as, in increasing order of engagement:

- perceiving;
- pointing;
- accessing;
- annotating;
- editing;
- creating.

A higher level of shareability usually relies on the lower ones (i.e., users need to be able to access and edit shared content in order to create new content). Different types of technologies (e.g., single- vs. multi-touch interaction), combined with different interface design solutions (e.g., shared areas of the display real estate vs. personal territories of interaction) can affect the level of shareability of an interface. Our work explores ways of supporting collaborative creativity by designing socio-technical environments for it.

4 Designing environments for co-located collaborative creativity

A first step to enhance collaborative creativity is to acknowledge that the layout of the environment affects the perception of ideas. As discussed by Fischer [3], "Creativity occurs in the relationship between an individual and a society, and between an individual and his or her technical environment". In this respect, the design challenge lies in creating socio-technical environments which are suitable for supporting and stimulating such synergistic processes. From this perspective, the social, physical, and technological contexts acquire a main role in determining communication as well as cognitive processes.

One can then approach the design space by reflecting and building on some of the existing CSCW theories of collaboration, as well as on cognitive theories of the use of physical space for creative tasks. Mutual visibility of action is a main design principle for groupware, as it provides awareness of what other colleagues are doing and how the actions of group members affect the shared artifacts, thus enhancing group awareness [2]. *Group awareness* (i.e., the condition in which members perceive the presence of other group members and the possibility to communicate with them) provides chances for informal communication, which in turn can strengthen the ties of a group, as well as facilitate the transfer of essential information related to task-specific activities. In this respect, the size of a tabletop display supports group awareness as it allows participants to see each other's actions, movements, expressions, and gestures in real time. Large surfaces also enhance a sense of common ground by affording the display of resources, as well as of the results of participants' interactions on a single shared area. Furthermore, looking at the cognitive benefits of the physical space, large surfaces support epistemic actions [10] by enabling temporary spatial arrangements: These, in turn, enhance creative processes by allowing the externalization and visualization of different alternatives [9].

Starting from these considerations, one can then explore possible ways in which the embodiment of computer-mediated interaction capabilities in a physical and social setting can either:

- augment existing collaborative tasks. In this sense, some of the qualities of digital technologies (e.g., storage of data, easy replication and distribution of digital video, for example) can provide novel features to traditional processes, thus potentially enhancing them and/or their management (e.g., [6]);
- turn single tasks into collaborative ones: for example, take tasks that were previously performed alone on a personal workstation, such as video editing, and have them performed collaboratively on a large interactive surface (e.g., [13]).

Our design explores both approaches.

5 Design goals

In order to continue the ad-hoc and spontaneous quality of the video capture process, we would like to support spontaneous collaborative creativity in the editing process too. Our design should therefore allow:

- spontaneous work,
- explorative work.

In order to achieve these higher level goals, we wanted our design to inherently motivate its users to explore by being playful and graspable.

Regarding the second approach, i.e., turning a single user task into a collaborative one, our design aimed at providing an interaction technique which could support directness and control, thus enhancing body language and group awareness. Video editing by single users is nowadays mostly done on PCs following the WIMP interaction paradigm. For analogue film the work is still often done on a film editing table. Besides the completely different concepts and a number of technical differences, the modality of interaction concerning the physical engagement is intrinsically different too. The analogue editing offers a direct interaction with the material and gives the editor the feeling of direct control. The digital editing, using a desktop PC, looses this feeling. With large interactive surfaces, direct interaction techniques become available, which could give back the feeling of direct control. Such a consideration influenced our design.

The functions needed for video editing can be divided into relatively simple and more complex functions. The simple functions particularly support the easy, explorative, and playful way of editing. The more complex functions are often optional, and do not profit quite as much from fluid collaboration as the simple ones. Therefore, we gave priority to not disrupting collaborative creativity with too complex functions. If required, complex functions can always be execute after the creative process itself, in a subsequent fine tuning phase.

The simple functions provided by our design are:

- sharing clips for collaborative use;
- splicing several media clips to a sequence;
- splitting sequences or clips;
- sharing a preview of clips or sequences;
- playing, pausing, scrubbing;
- changing the speed of clips.

The more complex functions are:

- editing the volume by setting key frames and changing their values (e.g., fading the sound at the beginning or at the end);
- trimming clips regarding their exact in- and out-points in the range of few frames (fine cut).

Figure 1 shows a top view of the appliance with its most important design elements for all simple functions, to give an impression of our design choices. The different parts and their functions will be discussed in the design section.

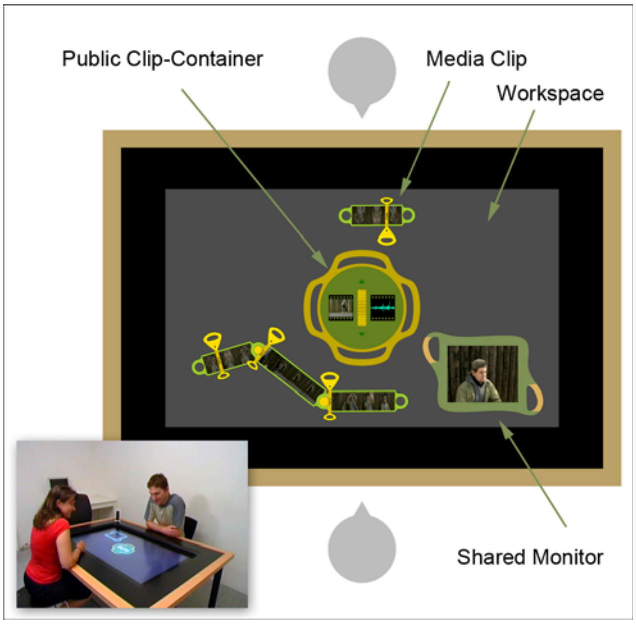

Figure 1. Arrangement of the components on the tabletop.

6 Related work

The Tangible Video Editor [15], the Moving Pictures [14] as well as the VideoPlay project [12] all use tangible interfaces for sequencing video clips. All three consider a multi-user environment and promote a playful experience for spontaneous and explorative work. In all these three examples video clips are represented by physical objects: Sequencing video clips means arranging and "playing" with these objects. This choice is motivated by the underlying goal to enhance a very explorative approach, thus affording a high potential for creativity. In our design we also aimed at supporting an explorative and playful approach, but modelled the interaction objects in 2D for direct touch interaction. The look and feel resembles real objects, which are graphically represented. At the cost of a truly haptic quality, this provides substantially more freedom for other design decisions: Displayed clips can actually adapt their size to the clip duration, which physical blocks obviously can't. Also, physical interface objects can't be cut directly like film strips. The virtual film strips we designed can be cut and pasted arbitrarily. A fully digital display of the interface objects also makes them persistent. A project created with our prototype can be stored for later use and further modification. Similar to related work, though, we introduced a physical tool for accomplishing more complex functionalities of video editing (cf. Section 7, Figure 8). In our case, this tool had some specific editing functionality: This was motivated by the idea of exploiting the manipulation vocabulary of a physical object, which can assume different functions when orientated differently, for example. More details are presented in the next section.

7 Implementation and design of the interface

Our prototype was implemented to run on a touch-sensitive table in our lab (see Figure 1, left corner). The table frames a NEC LCD display with a resolution of 1366x768 pixels. A SmartTech Actalyst touch panel uses four infrared cameras, placed in the corners of the frame, to track two fingers simultaneously. To recognize the marker of the tool a further greyscale firewire camera captures a top view of the table. The prototype of the interface was implemented in C# by using the Microsoft .Net-Framwork. The core of the application deals with the editing functionality and builds on the DirectShow-Editing-Services High-Level-API. To achieve our design goals we needed a way to visualize virtual objects which are looking like real bricks or gaming pieces. This visualization had to be fast enough for a real-time interaction. Therefore we decided to design all the visual objects in a 3D-program and used the managed DirectX version of Microsoft's Direct3D-API for the real-time visualization.

In terms of the user interface, here are the design decisions we made according to the goals and the interaction functions described above:

Sharing Media. To make them available for common use, all media clips (which can be either video or audio) from all group members are collected in a shared area called Public Clip-Container (see Figure 2). The Public Clip-Container is movable and, as shown in Figure 1, it initially rests in the center of the table for equal access by all users. An automated rotation of the area considers the position of the users around the table. Depending on the handle by which the container is grabbed, the preview frames of the clips are rotated towards the user. This assumes that the user would use the handle closest to her/him. Just as the real life situation by which it was inspired (Figure 2, left), the Public Clip-Container is constructed for two-handed interaction.

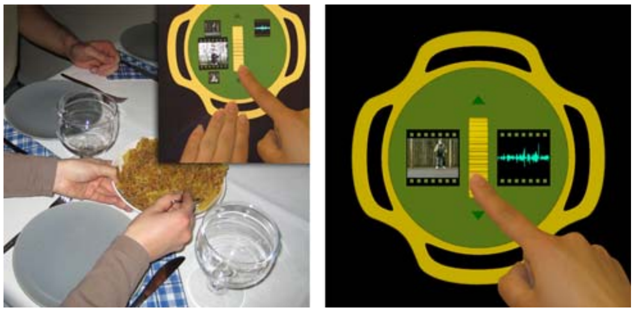

Figure 2. The Public Clip-Container gives access to all media clips and is built on the metaphor of a central plate for food.

Whoever wants to use it has to pull it by the nearest handle with one hand and interact with it by using the other hand. Within the Public Clip-Container, every clip is represented by a single poster frame. Video clips are on the left and audio clips are on the right. Using the wheel in the middle (see Figure 2, right) lets the user browse through the clips by rotating them on a virtual cycle about the central wheel.

Figure 3. Shared Monitor on the left, gesture for sharing on the right

The Shared Monitor for *previewing all media clips or sequences* is shown in Figure 3. It is for public use and therefore movable and re-orientable. It has two handles for interaction. Using one of the handles, the monitor can be moved around the table. The handles also recognize a flicking-like gesture for passing the monitor to another person. If this gesture is used on one of the handles, the monitor automatically moves in the respective direction. The speed and distance of movement is controlled by the force of the gesture, and a simulated friction eventually stops the movement. Using both handles simultaneously with two hands, the monitor can be scaled and rotated for maximum flexibility.

Splicing clips to a sequence is done on the workspace as shown in Figure 1. The workspace lets users create a common project by placing clips on it. A media clip, represented by a poster frame in the Public Clip-Container, can be placed there by just dragging it onto the workspace.

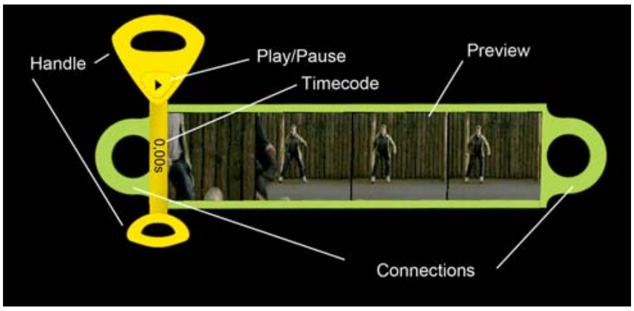

Figure 4. A video clip which is placed in the workspace.

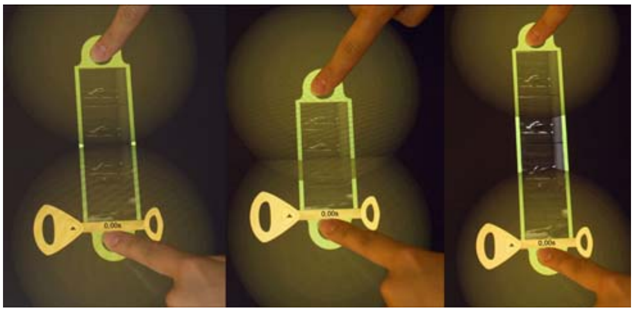

Figure 6. Changing the speed of a clip.

When the clip is dropped, it grows into a virtual film strip as shown in Figure 4. The size of the clip is now related to its duration. It also has two connections, one at the beginning and one at the end. With these two connections, clips can be spliced much like the mechanical coupling of toy trains. The visual feedback for spliced clips is the yellow coupling as shown in Figure 5 on the right. In this way, the workspace doesn't distinguish between areas of personal work and group work. Users are flexible to work where there is enough room – much like on a real table.

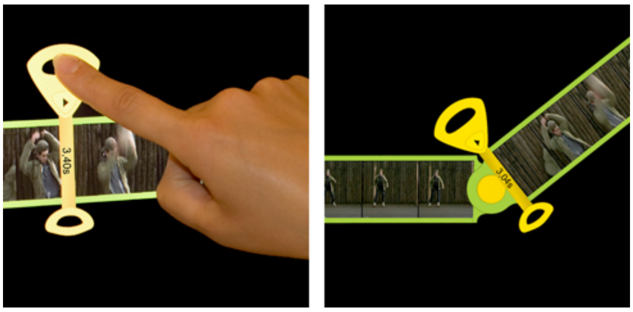

Figure 5. Scrubbing through a clip on the left, and two clips spliced to a sequence on the right.

Play, Pause and Scrub are functions which affect the preview in the shared monitor. The clip has a controller alluding to the playing head on a tape, which controls the actual frame to be displayed. The yellow controller has two handles for moving it along the clip. When the controller is moved, the frame at this position is displayed on the shared monitor. A Play/Pause button on the controller starts an autonomous movement of it, resulting in a real time playback of the clip. When two or more clips are connected to a clip chain, the movement – and with it the playback – will continue until the end of the last clip in the chain. When playback is running, the play button changes to a pause button.

Splitting sequences or clips are important functions to support the explorative goal of the design. The connection between two spliced clips (see Figure 5, on the right) can be released by pressing the yellow coupling.

The two handles of the yellow controller imply a two-handed interaction, which is in fact implemented. If the controller is moved to a position at which the user wants to cut the clip, he or she can do this by using the two handles like a pair of scissors (as shown in Figure 7).

Changing the speed of a clip is another two-handed interaction with a clip. The clip can be stretched or compressed by touching its two connectors (Figure 6) and adjusting their distance. Two circles around the clip connections provide additional cues: If they barely touch as shown on the left, the clip will be played in normal speed. An overlapping of the spheres indicates a faster playback and a space between spheres indicates slow-motion playback. As a side effect, this interaction technique also allows to rotate the clip.

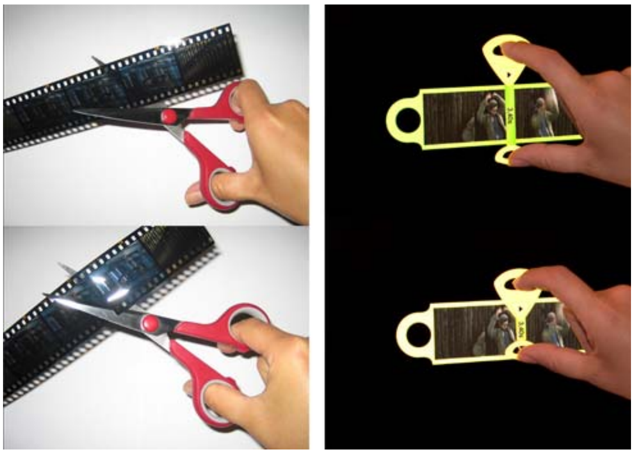

Figure 7. Cutting a clip on the right and its physical metaphor on the left.

The rotation of the clips is part of the free arrangement of the common project as shown in Figure 1. This allows a visualization of the cutting rhythm or simply to use all available workspace. The concept of the free arrangement provokes creativity by its lack of restrictions.

In the design goals section, we stated that we didn't want to interrupt collaborative creativity with too complex functions. In order to achieve this, we designed a hybrid (physical + digital) tool. It can be fetched and used on the tabletop surface only as required and thereby reduces the complexity when it is absent.

The Hybrid Tool consists of one physical and two digital components [1]. The physical component is thin and cylindrical and has two differently marked sides. Each side has a special function and an associated digital part

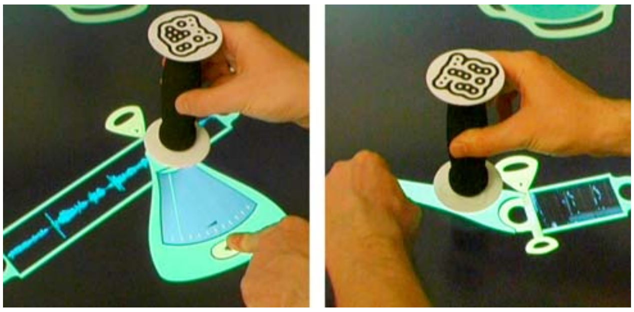

Figure 8. Hybrid Tool with different functions, depending on the side of the physical component.

(displayed as graphics on the tabletop) to enhance this. The Hybrid Tool and its use are shown in Figure 8. The tool is put down on the table only when the complex functions provided by it are needed. The functions depend on the position of the tool. It may affect either a clip or just a frame in the clip on which the tool is placed. After the tool is placed at the desired position, its functions are controlled by rotating it, much like a volume knob on a radio. The functions it executes have the character of precise actions. The use of a tool like this mediates more precision, because it is a smooth and well known direct interaction and alludes to using a pen or a screw driver.

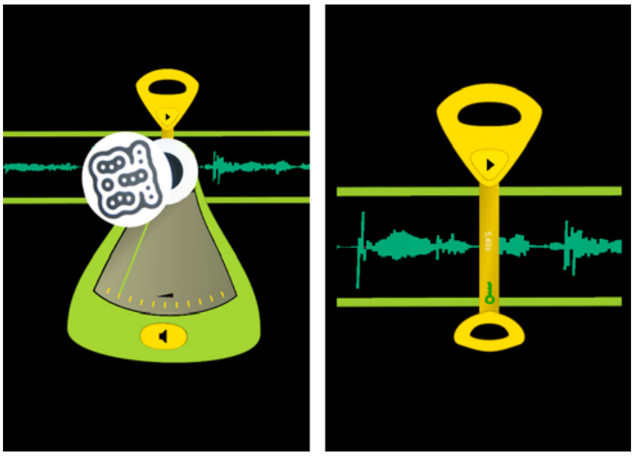

Figure 9. Changing the volume by setting a keyframe

Changing the volume is done by means of the Hybrid Tool. When its physical component is placed on an audio clip, the digital component appears (Figure 9, left). Pressing the button on the digital component activates the function at the current position. When the button is held and the physical component is rotated, this affects the value of the key frame, effectively changing the audio volume at this position. The volume level is indicated by a volume meter. When the button is released, a key frame is set and indicated by a small icon (Figure 9, right).

Trimming clips is also done with the Hybrid Tool, but with its other side. Trimming is necessary, when the rough cut is ready and only small adjustments in the range of only a few frames are needed. For this, the tool is used at the start or end connector of a clip like a combination wrench (Figure 10). The interaction technique is similar to the volume component. When the physical component is put down on a clip connector, the digital component appears (Figure 10, right). Pressing the button initiates the function and rotation trims the clip.

Figure 10. Trimming the cutting point of a clip on the right based on the physical metaphor on the left.

8 Evaluation and lessons learned

The appliance was informally evaluated in a focus group with 5 participants, consisting of 3 female and 2 male subjects at the age of 16 years. All participants owned a mobile phone with a camera, and we asked them about their practice with video capture and video work. Confirming the results of other research [8], their use of the camera and video capture is spontaneous, and they do not normally edit the clips they take. They all mentioned that they never use Bluetooth because they found it too complicated, and they normally mostly store clips on their mobile phone only. Some said that they occasionally download clips to their computer to share them on YouTube if they think they are funny. When we asked whether they could imagine wanting to edit the videos, they mentioned that it is too complicated for them and that it required too much effort for the quality. However, they could imagine adding texts if it was made easy enough.

After this preliminary discussion, the participants were introduced to our collaborative video editing scenario. The video of the Microsoft Surface™ [11] was presented to them in order to show the potential of using a tabletop technology in a collaborative situation. After this, they were introduced to the functionalities and interaction techniques of our appliance, and then they were asked to interact with the appliance we designed. Considering the technical limitations of the prototype, which allows for simultaneous input in two points only, it was not possible to conduct a controlled test in which multiple users could actually simultaneously edit the video. Our main focus, though, was on exploring the scenario and discussing with the participants, whether they could imagine using such an appliance, their concerns and requirements, and whether our design choices had met our goals in terms of design.

Considering the interface, the participants showed no difficulty in performing most of the functionalities (i.e., selecting clips from the Public Clip Container, changing speed, connecting and trimming clips, and sharing the monitor). The general concept of arranging the clips as a free form chain and the two-handed interactions for changing speed and scaling the monitor were particularly appreciated. Interestingly, on the other hand, they all had difficulties in interacting with the physical tool for editing the audio files and changing in- and out-point.

Concerning the scenario, instead, they stated that they would be very motivated to interact with the appliance, especially in a café or public space, for sharing and creating holiday or party movies, for example. On the other hand, they mentioned that their engagement would depend on how easy it was to bring their digital videos from their mobile phones to such a tabletop appliance, and then save it in a format that allows them to share the created videos in other contexts. Finally, the quality of the video seemed to be a main issue. The low quality of the video that can be captured by most of the mobile phones currently available (and owned by teenagers) is mostly too low for them to pay off for the effort of editing the video clips.

9 Implications for design

The design and evaluation of our appliance suggest that there is a potential for novel collaborative and creative scenarios around videowork. For our target users, engaging directly with the appliance allowed them to consider such possibilities and to reflect upon their requirements for this kind of interactive environments. In this sense, the instantiation of our design goals and ideas into an experience prototype, although not completely functional, allowed us to convey our vision and elicit further requirements that can inform the design of such collaborative environments.

On the one hand, the direct control and continuous, two-handed interactions seemed to be particularly appreciated, thus suggesting that giving more “physicality” to the editing task makes the activity more engaging, and in turn motivating. On the other hand, the use of the physical tool appeared to be problematic. This might depend on the design of the physical element, which constitutes a new tool users need to learn how to handle; or, it might be due to the functionalities to which it is associated (i.e., editing of in-point and sound volume in a specific key frame), which are more complex and are not necessarily performed in basic video editing.

Making the editing activity playful and collaborative seemed to meet their goal for shareability and entertainment: It is interesting to note, in this respect, that they could imagine such an appliance in a public space, where the spontaneity of the interaction is obviously enhanced.

Finally, the transition between interaction environments remains a main concern that the design of video technologies needs to address: How to transfer video material across displays is a main issue. If the effort required for such transitions is too high, this might hinder the attractiveness and motivation of collaborative scenarios.

Acknowledgements

This research was funded by the German Research Council (DFG). We would like to thank the participants of our focus group for their valuable feedback.

References

[1] A. Butz, O. Hilliges, L. Terrenghi, D. Baur, Hybrid Widgets on an Interactive Tabletop. In *Ubicomp ’07: Adjunct Proceedings, Demo session*, Innsbruck.

[2] P. Dourish, V. Bellotti, Awareness and coordination in shared workspaces. In *CSCW ’92: Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pp. 107–114. ACM Press.

[3] G. Fischer, Symmetry of ignorance, social creativity, and meta-design. In *C&C ’99: Proceedings of the 3rd conference on Creativity & cognition*, pp. 116–123. ACM Press.

[4] R. Gold, The Plenitude: Creativity, Innovation, and Making Stuff . The MIT Press, 2007.

[5] O. Hilliges, M. Wagner, L. Terrenghi, A. Butz, The Living-Room: Browsing, Organizing and Presenting Digital Image Collections in Interactive Environments. In *Proc. of IE’07,* Ulm, Germany.

[6] O. Hilliges, L. Terrenghi, S. Boring, D. Kim, H. Richter, A. Butz Designing for Collaborative Creative Problem Solving. In *C&C ’07: Proc. of the 6th International Conference on Creativity & Cognition*, pp. 137–146. ACM Press.

[7] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner The reactable: Exploring the Synergy between Live Music Performance and Tabletop Tangible Interfaces. In *Proc. of TEI ’07*, pp. 139–146. ACM Press.

[8] D. Kirk, A. Sellen, R. Harper, K. Wood, Understanding Videowork. In *Proc. of CHI '07,* pp. 61-70. ACM Press.

[9] D. Kirsh, The intelligent use of space. *Artificial Intelligence*, 73 (1-2): 31–68 (1995).

[10] D. Kirsh, P. P. Maglio,. On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4):513–549 (1994).

[11] Microsoft Surface http://www.microsoft.com/surface

[12] S., Taylor, S. Izadi, K. Ozenc, R. Harper, R. Banks, A. Butler VideoPlay: Playful and Social Editing of Video using Tangible Objects and Multi-touch. In *Proc. of IEEE TABLETOP ’08*.

[13] L. Terrenghi, T. Fritsche, A. Butz, The EnLighTable: Design of Affordances to Support Collaborative Creativity. In *Proc. of the International Symposium on Smart Graphics*, 2006, Springer Verlag.

[14] C. Vaucelle, D. Africano, G. Davenport, M. Wiberg and O. Fjellstrom, Moving pictures: Looking Out/Looking In. *Proc. of SIGGRAPH ‘05*. ACM.

[15] J. Zigelbaum, M. S. Horn, O. Shaer and R. J. Jacob, The Tangible Video Editor: Collaborative Video Editing with Active Tokens. In *Proc. of TEI '07*. ACM Press.