Toward More Efficient User Interfaces for Mobile Video Browsing: An In-Depth Exploration of the Design Space

Jochen Huber, Jürgen Steimle, Max Mühlhäuser
Telecooperation Group
Technische Universität Darmstadt
Darmstadt, Germany
{jhuber, steimle, max}@tk.informatik.tu-darmstadt.de

ABSTRACT
Increasingly powerful mobile devices enable users to access and watch videos in mobile settings. While some concepts for mobile video browsing have been presented, the field still lacks a general understanding of the design space and of the characteristics of interaction concepts. In order to improve user interfaces for mobile video browsing, this paper includes three contributions. First, we setup a design space for mobile video browsing and contribute seven novel interface concepts. They rely on GUI-based, on touch-gesture-based, and on physical interaction. Second, we present the results of an in-depth evaluation and comparison of these concepts. They are based on an ascriptive analysis of 18 hours of video observations from a controlled experiment with 44 participants. The results provide insights into common usability errors and misconceptions. Third, we derive implications for the design of mobile video browsers to minimize errors and to increase usability.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms
Design, Human Factors

Keywords
Video browsing, mobile device, user interface, evaluation, video player.

1. INTRODUCTION
Increasingly powerful mobile devices like Apple's iPhone are currently dramatically changing how we perceive multimedia when being on the move. Users are able to access a constantly increasing number of video streams almost anytime and anywhere. Videos are not only watched for entertainment during leisure time [13], but also used at work, e.g. for learning on the job or for mobile learning [7]. As mobile devices typically have recording capabilities, users can not only watch, but also record and share video data virtually anywhere.

Mobile video browsing differs from classical video browsing in a desktop setting in several aspects. On the one hand, mobile devices have severe restrictions due to their form factor (e.g. small displays), but on the other hand they also offer support for novel forms of input (e.g. direct touch and physical input). Moreover, users in mobile settings typically cannot devote their full attention to the user interface. Hence, even more than with desktop interfaces, high efficiency and effectiveness of the user interface is crucial [4].

Some approaches for novel and efficient user interfaces for mobile video browsing have been presented [10,17]. While these are valuable contributions for specific aspects of mobile video use, research is still very fragmentary, pointing on individual aspects of the design space. The field still lacks a general understanding of the design space for mobile video browsing and of the characteristics of specific interaction concepts. A precise knowledge of the advantages and pitfalls of different interface concepts, of frequent use patterns and of recurrent misconceptions related to these concepts is likely to significantly increase the quality of future mobile video browsers. To state only one example, our analysis shows that the video player included in Apple's iPhone has significant drawbacks and also points out how future versions could be designed to be more efficient to use.

In order to advance interaction with mobile video browsers, we have modeled the design space for mobile video browsing. This covers both GUI-inspired interaction concepts and more innovative concepts for mobile devices, such as gesture-based and tangible interaction. Within this design space, we have designed several novel interface concepts for mobile video browsing each incorporating different characteristics of the design space. The interfaces have been implemented for the Apple iPhone. An in-depth evaluation of 8 different interface concepts (7 novel concepts, one standard interface) allows a broad comparison and a deep understanding of their respective advantages and pitfalls. Based on an ascriptive analysis of more than 18 hours of video observations, we evaluated usability criteria and performed a systematic analysis of usability errors. This analysis is both more comprehensive and more detailed than prior work. Based on our findings, we present implications for the design of future mobile video browsers.

Hence, the contributions of this paper are

- novel interface concepts for mobile video browsing,
- an in-depth evaluation of these concepts and
- design guidelines for mobile video browsers.

The remainder of this paper is organized as follows. First, we illustrate common use patterns of mobile video browsing and describe the design space. We then introduce the interface concepts and present results of the usability and user experience evaluation. Moreover, we report the results of a comprehensive usability error analysis. Based upon these results, we outline design implications for future mobile video browsers. Thereafter, we discuss related work and conclude with an overview of potential future work.

2. USE PATTERNS OF MOBILE VIDEO BROWSING

In order to set-up the design space of mobile video browsing, it is important to understand the patterns of use. O'Hara et al. [13] studied how mobile devices are used for consuming videos for the purpose of entertainment. The study shows that patterns of mobile video consumption differ from just "watching TV" in a mobile settings. The authors present a range of deeper motivations related to the social setting. Mobile video browsers are used for sharing video contents by watching them with other people in social situations. In contrast, they are also used for being in proximity to others (e.g. with family at home) while not disturbing them with own content. Moreover, mobile video can also be characterized as a "privatizing technology": by watching video contents, users can withdraw themselves from public space when being in the proximity of other people, e.g. in public transport.

Mobile videos browsers are not only used for entertainment, but also for mobile learning and training on the job. The ubiquitous availability of multimedia learning material through services like iTunes U [8] or OpenCourseWare [14] has paved the way for groundbreaking changes in mobile learning. A recent study [7] found a shift in the usage habits of students towards using the mobile version of lecture recordings. These settings of learning and working result in different use patterns than entertainment.

Much more than entertaining videos, instructive videos are watched in a non-linear manner. Analogously to the work with textbooks, users tend to watch specific passages of interest and jumping between different passages instead of linearly consuming the entire video, particularly in case of lengthy lecture recordings.

Besides watching individual videos, the interrelationship of several videos (e.g. as hyperlinks in so-called *hypervideos*) is of major importance for successful learning processes and trainings. The relationships are crucial for contrasting and integrating knowledge which is contained in related videos. This can be compared to reading books and articles, where we follow references and compare and integrate information from various documents. For example, various topically related videos from different institutions allow learners to receive elaborate explanations for a certain problem and can be used to gain deeper insight into a specific problem domain from a slightly different point of view. This practice is possible nowadays due to the vast amount of video recordings available online from various institutions.

3. DESIGN SPACE

In the previous section, we have shown that patterns of mobile video browsing include using several inter-related videos. This has direct effects on the design space. Not only the efficient navigation within an individual video, but also within collections of several inter-related videos is crucial for video browsing, e.g. in knowledge work. These aspects require

- Getting detailed information on the current topic contained within an *individual* video *segment*,
- an efficient overview on a *large video* with quick and easy access to any of the contents and
- quick and easy navigation to information which is related to the current topic within a *collection of inter-related videos*.

These three levels characterize the complexity of navigation and therefore are one dimension of the design space. It is depicted as the vertical axis in Figure 1.

The second dimension of the design space characterizes the type of interaction used in the video browser. It is depicted as the horizontal axis in Fig. 1. These types range from classical interactions, which are well-known from desktop computing, to innovative interactions which leverage the additional capabilities of mobile devices, such as multi-touch displays and inertial sensors. Although the boundaries between the interface types are not selective, we regard them as discrete categories. This comprises

- classical *graphical user interfaces* ported to the small display of mobile devices,
- interfaces that rely on *gestures* performed by touching the display of the mobile device and
- interfaces that rely on manipulating the device itself in the *physical* space.

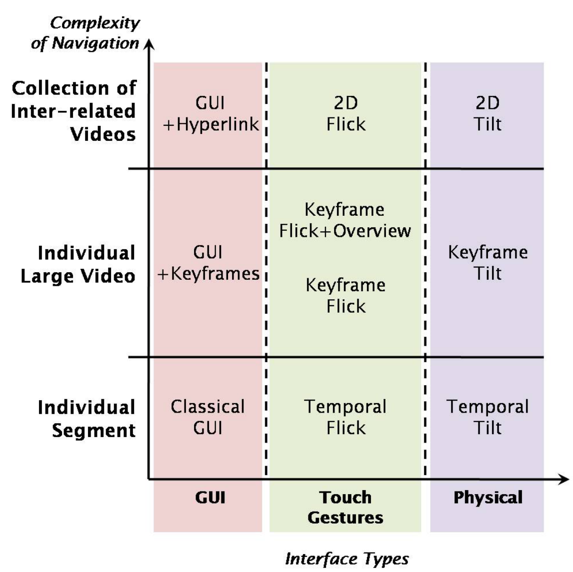


Figure 1. Design space for mobile video browsers

The combination of both dimensions allowed us to identify various interface concepts. These are presented in the following section.

4. INTERFACE CONCEPTS

The design space contains interface concepts with different affordances according to the complexity of navigation. We have taken one existing interface concept (the standard iPhone video player) and designed and implemented 7 novel interface concepts.

The interfaces addressing the navigation within a large video utilize keyframe abstraction [18] to allow for an efficient overview over and a quick access to the contents of the video. The interface concepts for the navigation within a collection of inter-related videos are based upon hyperlinks. These hyperlinks exist between semantic segments of the video (e.g. key frames of a video or slides of a presentation recording), which can point to video segments or even a complete video. The hyperlinks are *1:n* relationships. Hence one semantic segment can point to various other semantic segments. It is out of the scope of this paper how these links are created, since we focus on the navigation concepts. Hyperlinks could be created automatically through multimedia information retrieval [15]. Furthermore, the interfaces could be enhanced to allow users to manually create (and share) links between videos.

4.1 GUI Navigation

A first class of interface concepts is based on the traditional GUI metaphor.

The standard iPhone video player (see Fig. 2a) is a **classical GUI** for the navigation within individual video segments. The interface elements in the center feature a *play/pause* button and two buttons to jump to the next (or previous) file or video in the playlist. The latter buttons can also be used for *navigation* (rewind or fast-forward respectively), when being pushed for a longer period of time. The slider below the buttons allows controlling the volume. The timeline at the top is used for the navigation by dragging the knob along it. However, the timeline itself cannot be manipulated. By tapping onto the knob a little longer and dragging it vertically activates a technique called *scrubbing* (comparable to the MobileZoomSlider [17]): when a user drags the knob farther down vertically, the interface adapts the navigation granularity (in discrete levels). For instance by dragging the knob to the middle of the interface and then dragging it horizontally as if the user manipulated the timeline, the navigation speed is being reduced by 50% and the granularity is therefore increased.

Our **GUI+Keyframes** interface concept (see Fig. 2b) is an enhanced version of the iPhone video player. We have added two buttons, which allow switching back and forth between semantically segmented units and therefore *navigating* more quickly through a rather lengthy video. These buttons are located to the left (and right) of the previous (and next) buttons.

As a further enhancement of the GUI+Keyframes interface concept our **GUI+Hyperlink** interface concept (see Fig. 2c) supports the navigation between topically related videos, which are contained in an inter-related video collection. By tapping onto the screen, the user can access a list of hyperlinks relating the current segment to other videos. By tapping onto one of the links, the related video segment is replayed. The button located to the left of

c) GUI+Hyperlink interface for a collection of inter-related videos

b) GUI+Keyframes interface for an individual large video

a) GUI for an individual video segment

Figure 2. GUI-based interfaces

the list allows browsing back in the history, comparable to a back button in a web browser.

4.2 Gesture-based Navigation

The second class of interface concepts draws more extensively on the direct touch input capabilities of modern mobile devices. Instead of buttons (like in traditional GUIs), we leverage flick gestures that can be easily performed by touching the display. The gestures are inspired by the analogy to thumbing through a book.

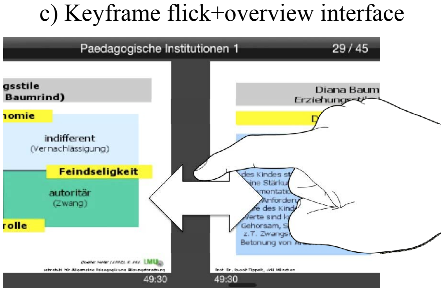
Our **Temporal Flick** interface concept is shown in Figure 3a. The timeline at the bottom of the interface is used for the *navigation* within an individual video. Additionally, users can navigate through the video's frames by flicking horizontally. Flicking from right to left fast-forwards and flicking from left to right rewinds, respectively. The cording speed depends on the length of the flick. Tapping the display toggles between the *play* and *pause* modes.

Two interfaces support the navigation within a large video. The **Keyframe Flick** interface (see Fig. 3b) is an extension of the temporal flick interface. When the video is *paused*, the current key frame is displayed and by flicking horizontally, users can switch back and forth between key frames. By tapping the display again, playback of the video is resumed.

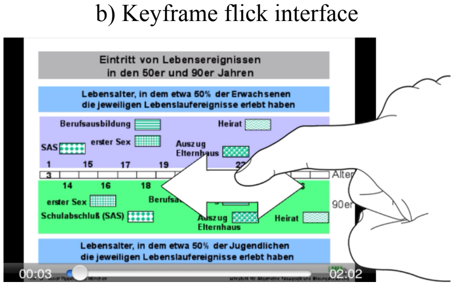
The **Keyframe Flick+Overview** interface (see Fig. 3c) extends the keyframe flick interface by an overview with thumbnails of all keyframes of the video. These are displayed in the lower part of the interface in a grid layout. The currently active key frame is highlighted. In addition to the flick interactions described above, the user can directly navigate to any key frame by tapping onto its thumbnail. Moreover, key frames can be skimmed very quickly by sliding the finger over the grid. Either rotating the device into landscape mode or double tapping on the current video in the upper part activates the keyframe flick interface.

For the navigation between topically related videos, we contribute the **2D Flick** interface concept (see Fig. 4a). We aim at providing an intuitive interaction technique, which allows users to follow hyperlinks and navigate easily within the navigation history. The major challenge is to prevent users from getting lost in too

c) Keyframe flick+overview interface



b) Keyframe flick interface



a) Temporal flick interface

Figure 3. Gesture-based Interfaces

much information presented on a small screen. Lost in Hypertext [3] is a well-known phenomenon, which may occur particularly in this situation. Therefore, we apply a spatial navigation concept: While segments within one single video can be accessed by flicking left and right (as described above), hyperlinks between different videos can be followed by flicking up and down.

Whenever a hyperlink is available, this is indicated by a small arrow in the upper right corner of the user interface (see Fig. 3c and 4a). When the user flicks downwards, the interface is being scrolled downwards, revealing related videos as shown in the lower interface screenshot in Figure 4a. In this example, two interlinked videos (visualized using grey boxes) contain relevant material. By tapping on one of the videos, the interface is being scrolled down further, displaying the interlinked key frames of the related video (see the upper interface screenshot in Fig. 4a). These can also contain topical relations to other videos, which are again visualized with a small arrow in the upper right corner.

By aligning semantically related videos vertically, the browsing history results in a vertical stack. This can be navigated by simply flicking vertically up and down respectively. Alternatively, to avoid repetitive flicking and to gain an overview on the browsing history, a visualization thereof can also be used for the vertical navigation as shown in Figure 4b. It is activated by tapping anywhere on the screen for more than one second and then displayed as an image on top of the current video. The visualization can be navigated by moving the finger vertically across the images.

a)

b)

Figure 4. a) Vertical navigation between videos, b) Visualized browsing history

4.3 Physical Navigation

A third class of interface concepts leverages the affordance of mobile devices to be manipulated in the physical space. Our *Temporal Tilt* user interface (see Fig. 5a) contains only one visi-

b) Usage of the tilt interface, arrows indicate tilt directions

a) Temporal tilt interface: illustrating the interface elements

Figure 5. Physical interfaces

ble interface element: the circular timeline at the bottom right. Tapping the display anywhere toggles between the *play* and *pause* modes. When tapping the display for a longer period of time, a slider appears (see top right in Fig. 5a and 5b near thumb) that can be dragged to the right to activate the *navigation* mode. By tilting the device to the right, the user can fast-forward the video. By tilting to the left, the user can rewind the video. The interface features two discrete cording speeds, depending on the tilt angle (the greater, the faster or slower respectively). For detecting the tilt action, the device's accelerometer is used.

The same tilt-based concept can also be used for navigating between segments of large videos (**Keyframe Tilt**). Instead of fast-forwinding or rewinding the video, tilt actions result in jumping forth and back between segments. The **2D Tilt** concept involves not only tilting to the left and to the right, but also tilting upwards and downwards for navigating within inter-related video collections, similarly to the *2D Flick* interface. For our study we have only implemented and evaluated the temporal tilt interface for reasons that will be discussed in the next section.

5. USABILITY AND USER EXPERIENCE

The concepts described above have been evaluated in a controlled experiment. We evaluated the usability (focusing on efficiency, effectiveness, learnability, as well as user satisfaction) and user experience of each user interface. Moreover, we analyzed the specific advantages and drawbacks of each interface type for mobile video browsing in the design space, depending on the navigation complexity.

In the following subsections, we first describe our methodology and then report and discuss the results of the usability and user experience evaluation.

5.1 Methodology

We have conducted a controlled experiment with 44 participants (30 male, 14 female) from different scientific backgrounds (i.a. mathematics, social sciences, medicine, pedagogy, physics and design). Each single-user session had a duration of 2 hours. The tasks of the participants comprised simple fact finding and more complex knowledge integration tasks. For each task, a different data set was utilized to exclude any learning effects. Moreover, the order in which the interfaces were presented to the participants was counter-balanced. The sessions were video-recorded, relevant completion times were measured and semi-structured interviews were conducted. Additionally, quantitative feedback was gathered using the well-known standard usability scale (SUS) questionnaire [1]. The attractiveness of each user interface has been assessed using the AttrakDiff questionnaire [6]. Our analysis comprises statistical measures and a detailed ascriptive analysis of 18 hours of video recordings.

5.1.1 Navigation in an Individual Segment

In order to evaluate and to compare the interface concepts for navigation within an individual segment, the participants were asked to perform two different fact-finding tasks with each user interface. As data, we utilized videos of about 5 minutes length. The first task required textual orientation, whereas the second task focused on visual orientation, since the user's orientation is crucial to quickly retrieve a desired part of a video. The participants had to fulfill the following tasks:

- **Task 1:** The participants were asked to search a short video for a certain topic. The position within the video was not revealed to them beforehand.
- **Task 2:** The participants had to find a specific scene in the video. To support navigation, they were shown a distinctive key frame of the scene beforehand.

5.1.2 Navigation in a Large Video

To assess the navigation in a large video, the participants were asked to complete three different fact-finding tasks. As data, we used lecture recordings of each about 90 minutes length and the corresponding slides as key frames. The tasks required visual orientation within a video (task 1 and 3), as well as textual orientation (task 2):

- **Task 1:** The participants had to search a given slide within a lecture recording without prior knowledge of the lecture.
- **Task 2:** The participants were asked to find a certain topic. They were advised of the fact, that it was contained in the last third of the lecture.
- **Task 3:** The participants had to navigate to the slide which directly follows the one found in the first task.

5.1.3 Navigation between Inter-related Videos

We also assessed the navigation in a collection of inter-related videos. The collection consisted of 7 lecture recordings (each about 90 minutes) and 6 news broadcasts (each about 15-30 minutes). We segmented the videos and manually related the segments topically. The participants had to fulfill the following tasks:

- **Task 1:** The participants were asked to complete a complex visual and textual fact-finding task involving multiple videos using both interfaces.
- **Task 2:** The participants had to complete a knowledge integration task for a given topic covered in multiple videos.

In Task 2, we used the same data set for both interfaces. To exclude any learning effects, we used a between-subject design for this particular task.

5.2 Results

5.2.1 Navigation in an Individual Segment

Figure 6 shows an overview of the average time required for performing the tasks with each interface. Although the participants performed the tasks faster using the *temporal flick* interface, the one-way repeated measures ANOVA test revealed that the speed-up was not significant (see Table 1). The SUS score for the *classical GUI* browser is 76.53 (SD=15.01), 74.89 (SD=16.91) for the *temporal flick* browser and 62.44 (SD=16.87) for the *temporal tilt* browser. Hence, both the *classical GUI* browser and the *temporal flick* browser were perceived as most usable. The *temporal flick* browser was perceived as most attractive (with an average score of 4), in contrast to both the *classical GUI* and the *temporal tilt* browser (both with an average score of 3.5 on a 7-point Likert scale).

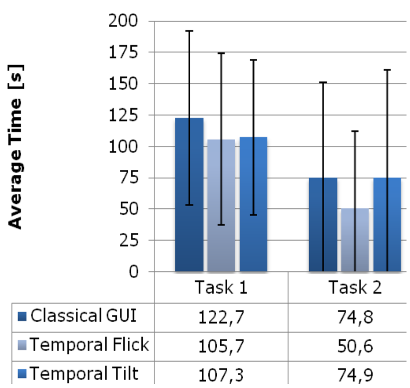


Figure 6. Average times for navigation in an individual segment (here and in the following error bars indicate standard deviation)

Overall, *temporal flick* turned out to be the best technique for temporal navigation within individual video segments. With the *temporal tilt* interface, users performed faster or not slower than with the *classical GUI*. We introduced the scrubbing feature of the *classical GUI* browser to our participants. However, none of the participants actually used this feature while browsing an individual segment.

Early focus group studies (which we have conducted after the implementation of the prototypes and before the actual evaluation), the qualitative analysis of the video data and statements from the semi-structured interviews all show that the tilt interaction is not adequate for continuously browsing within a video. Due to the larger viewing angle caused by tilting, the video on the display was less well viewable. Users felt that they need better

Table 1. ANOVA results for the navigation time in an individual segment and a large video

Navigation Complexity	Task	F	df	Sig.
Individual Segment	1	0.99	2, 86	> 0.05
	2	1.30	2, 86	> 0.05
Large Video	1	160.88	1.14, 49.09	< 0.001
	2	20.51	1.66, 71.37	< 0.001
	3	70.79	2, 86	< 0.001

visibility of the video for fine control in browsing. Moreover, users complained about the lack of haptical feedback. Vibrotactile feedback was not an option at the point of the study, since the iPhone as our implementation platform supports only one constant vibration force. Future work should examine more deeply how haptic feedback can improve physical interfaces for mobile video browsing. Because of these negative results even in this rather simple setting of navigation within an individual video segment, we have opted for not evaluating this concept for the more complex tasks.

An interesting observation we have made was that in spite of the rather continuous interaction techniques like flicking or tilting, which require continuous motion, users performed compound interactions, consisting of several discrete, additive flicks or tilts, e.g. flicking once for navigating 5 seconds forward, twice for 10 seconds, three times for 15 seconds and so on. Users also transferred interactions from everyday actions to the physical interface types. One particular user for instance drew circles onto the display using the *temporal flick* interface. He wanted to replay a certain scene in a loop. This underlines the potential of these novel interaction techniques for mobile video browsing.

5.2.2 Navigation in a Large Video

Figure 7 shows an overview of the average time required for performing the navigation tasks in a large video with each interface. The participants were able to complete all three tasks significantly faster using either the *keyframe flick* or the *keyframe flick+overview* interface than using the *GUI+keyframe* interface (see Table 1 for the ANOVA results and Table 2 for the Bonferroni post-hoc test results). Comparing the *keyframe flick* with the *keyframe flick+overview* interface, we found that the participants

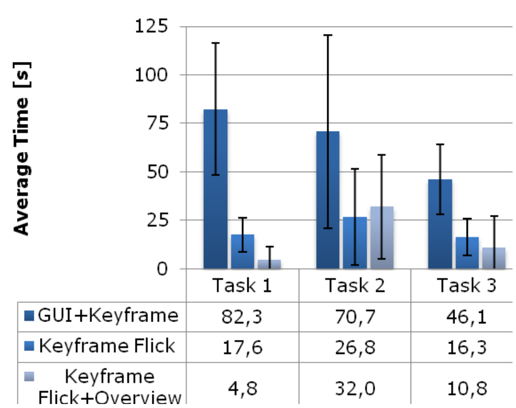


Figure 7. Average times for navigation in a large video

Table 2. Bonferroni test for the navigation in a large video

Task	Interface A	Interface B	CI.999 (lower)	CI.999 (upper)	Sig.
1	GUI+keyfr.	Keyfr. flick	42.85	86.51	< 0.001
	GUI+keyfr.	Keyfr. f+o	56.04	98.82	< 0.001
	Keyfr. flick	Keyfr. f+o	6.156	19.35	< 0.001
2	GUI+keyfr.	Keyfr. flick	10.37	77.45	< 0.001
	GUI+keyfr.	Keyfr. f+o	7.13	70.37	< 0.001
	Keyfr. flick	Keyfr. f+o	-26.08	15.77	> 0.05
3	GUI+keyfr.	Keyfr. flick	17.00	42.54	< 0.001
	GUI+keyfr.	Keyfr. f+o	21.34	49.34	< 0.001
	Keyfr. flick	Keyfr. f+o	-4.74	15.88	< 0.001

were significantly faster using the *keyframe flick+overview* interface for task 1 and task 3. The difference in task 2 was not significant. This is in line with qualitative findings from the semi-structured interviews. The participants stated that the *keyframe flick+overview* interface supports their visual orientation and navigation (as in task 1 and 3), whereas they prefer to skim through the slides by flicking horizontally when they have no visual clues (as in task 2).

The SUS score is 89.26 (SD=9.2) for the keyframe flick+overview interface, 90.34 (SD=7.48) for the keyframe flick interface and 61.59 (SD=16.1) for the GUI+keyframe interface. Hence, both keyframe flick and keyframe flick+overview interfaces were perceived as far more usable than the GUI+keyframe interface. Moreover, they were perceived as far more attractive (with an average score of 5 and 6 respectively) than the GUI+keyframe interface (with a score of 2.5 on a 7-point Likert scale).

5.2.3 Navigation between Inter-related Videos

In both tasks, a t-test (repeated measures for task 1 and independent measures for task 2 respectively) showed that the participants were significantly faster ($p < 0.001$) using the *2D flick* interface as shown in Figure 8. Moreover, statements in the interviews showed that the two dimensional browsing metaphor fosters the users' awareness of interrelated videos. Together, the above results show that a two-dimensional navigation metaphor supports the user's orientation when navigating across multiple videos.

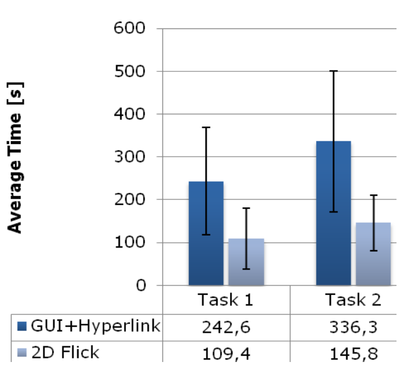


Figure 8. Average times for navigation in a collection of inter-related videos

The SUS score for the *2D flick* interface is 83.07 (SD=12.33), whereas the *GUI+Hyperlink* interface scored 58.98 (SD=19.76). Consequently, the 2D flick interface was perceived as far more usable than the *GUI+Hyperlink* interface. The *2D flick* interface was perceived as far more attractive with an average score of 6 than the *GUI+Hyperlink* interface with an average score of 3.5 on a 7-point Likert scale. In the interviews, the participants commented on the spatial concept as "clearly laid out" and they remarked that the vertical alignment of the related videos intensifies the visual relationship between the videos.

6. USABILITY ERROR ANALYSIS

For a deep understanding of the problems related to specific interface concepts, it is important to examine which errors are made when using the interfaces. We therefore performed a detailed analysis of usability errors which occurred during our experiment. From the detailed descriptions of individual errors, we derived a general error taxonomy for mobile video browsers. In the following subsections, we first report on our methodology, present our error classification and report the results of our error analysis. Together with our findings from the previous section, these provide the basis for design implications, which we will eventually derive in section 7.

6.1 Methodology

Inspired by [11], we utilize ascription for the analysis of the video data collected in the controlled experiment. Since we recorded the interactions from behind the participants' shoulders, the interfaces were always clearly visible. We have coded potential errors using a template describing the type of error, a detailed description of the error, its impact on the efficiency (e.g. loss of internal locus of control) and the occurrences per task and user interface. In the following, we outline the classes of our error taxonomy and then present the results of our analysis.

6.2 Results

6.2.1 Error Classes

We have identified the following four abstract error classes:

- **E1, Interface element not accessible:** The interface element could not be manipulated by the user. Typical reasons for this error are small or misplaced interface elements.
- **E2, Interface element was used incorrectly:** This type of error designates incorrect interactions. Common errors of this type are for instance wrong gestures.
- **E3, Interface elements misinterpreted:** Errors of this type mostly happened due to misconceptions. For instance, interface elements of the same type (e.g. two sliders), which were mapped onto different functions, were confused.
- **Slips:** All other errors, which do not belong to any of the above classes, are called slips (in the sense of [12]), e.g. users performed certain actions accidently.

The interface-specific errors within these categories are discussed in the following subsections.

6.2.2 Navigation in an Individual Segment

Although the GUI concept is the most well known interface concept in our design space, the participants committed most of the errors using the standard iPhone video player. They committed significantly less errors using both the *temporal flick* (73% less) and *tilt* browser (79% less). Other differences were not significant (cf. Table 4 for the ANOVA results and Table 5 for the Bonferroni post hoc test results).

Table 3. Amount of errors for the navigation in an individual segment (here and in the following, bold numbers indicate the peak per task)

	Class. GUI		Temp. Flick		Temp. Tilt	
Task	1	2	1	2	1	2
E1	**71**	**58**	**14**	4	0	0
E2	15	4	3	0	**13**	3
E3	5	4	5	10	7	6
Slip	31	34	11	**11**	11	**6**
Sum	122	100	33	25	31	15

The majority of the errors made with the *classical GUI* concept (the standard iPhone video player) were of type E1 (see Table 3). Users were unable to navigate through the video using the timeline placed at the top of the interface. Placing the timeline at the top of the interface causes severe issues. Most commonly, a mobile device is used in landscape mode to browse a video, since it offers the most screen real estate. In our experiment, a significant amount of participants held the device in both hands, such that only the thumbs are able to interact with the interface. The rest of the hand is located *behind* the device. Consequently, the interaction is highly limited by the length of the users' thumbs. Figure 9 shows one of our participants while trying to use his thumb to interact with the timeline of the iPhone video player. Since the timeline is located at the top of the interface and is not in reach for his thumb, he needs to lift his right hand, therefore occluding nearly the entire display real estate. In this case, the timeline should be placed at the bottom of the interface to (1) minimize the navigation paths and (2) prevent users from occluding the screen while using the timeline for navigation.

Slips were also a severe problem in case of the standard iPhone video player. Users often tapped onto the "next title" button accidently and therefore stopped the playback of the current video. Consequently, they lost the internal locus of control, had to restart the video and continue their video search from the beginning. Another difficulty with the iPhone video player was the fact that the same slider interface element was used for both timeline and volume control. Both interface elements got confused frequently (see E2 in Table 3).

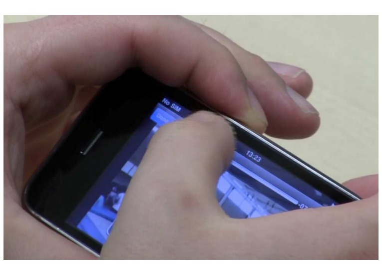


Figure 9. Timeline of the GUI interface is difficult to reach using a thumb, since it is placed at top.

Table 4. ANOVA results for the errors during the navigation in an individual segment and a large video

Navigation Complexity	Task	F	df	Sig.
Individual Segment	1	7.15	1.11, 47.90	< 0.01
	2	5.14	1.25, 53.82	< 0.05
Large Video	1	9.11	1.27, 54.77	< 0.01
	2	2.72	1.27, 54.74	< 0.05
	3	4.24	1.05, 44.93	< 0.05

Table 5. Bonferroni test results for the errors during the navigation in an individual segment

Task	Interface A	Interface B	$CI_{.95}$ (lower)	$CI_{.95}$ (upper)	Sig.
1	Class. GUI	Temp. Flick	0.10	4.03	< 0.05
	Class. GUI	Temp. Tilt	0.26	4.02	< 0.05
	Temp. Flick	Temp. Tilt	-0.47	0.60	> 0.05
2	Class. GUI	Temp. Flick	0.12	3.75	< 0.05
	Class. GUI	Temp. Tilt	-0.35	3.58	< 0.05
	Temp. Flick	Temp. Tilt	-1.10	0.47	> 0.05

Slips were the most dominant error type for both *temporal flick* and *tilt interfaces*. However, although they committed only little slips, the amount can be further reduced when users become more familiar with such novel interaction techniques. Regarding the temporal tilt interface, users had difficulties with enabling the navigation mode (error E2 in Table 3). Moreover, errors of type E3 were also problematic for both interfaces. With both interfaces it occurred that participants confused the correct navigation directions, e.g. flicking from left to right to navigate forward, instead of flicking from right to left. This is possibly due to differently remembered experiences and therefore a different mental model of the interface.

6.2.3 Navigation in a Large Video

Table 6 shows an overview of the amount of errors committed with each user interface. Virtually no errors were made with the gesture-based interfaces due to their high usability. Again, the ANOVA test (cf. Table 4 and Table 7 for the Bonferroni post- hoc test results) showed, that the participants committed significantly

Table 6. Amount of errors for the navigation in a large video

	GUI+Keyframe			Keyframe Flick			Keyframe Flick+Overview		
Task	1	2	3	1	2	3	1	2	3
E1	**22**	**32**	**54**	0	0	0	0	0	0
E2	6	8	7	**10**	**9**	5	0	6	0
E3	4	3	1	0	0	0	0	0	0
Slip	6	10	10	3	7	**10**	**1**	**6**	**1**
Sum	38	53	72	13	16	15	1	12	1

Table 7. Bonferroni test for the errors during the navigation in a large video

Task	Interface A	Interface B	$CI_{.95}$ (lower)	$CI_{.95}$ (upper)	Sig.
1	GUI+keyfr.	Keyfr. flick	-0.18	1.15	< 0.05
	GUI+keyfr.	Keyfr. f+o	0.25	1.43	< 0.01
	Keyfr. flick	Keyfr. f+o	0.03	0.52	< 0.05
2	GUI+keyfr.	Keyfr. flick	-0.45	2.13	< 0.05
	GUI+keyfr.	Keyfr. f+o	-0.35	2.21	< 0.05
	Keyfr. flick	Keyfr. f+o	-0.45	0.63	> 0.05
3	GUI+keyfr.	Keyfr. flick	-0.51	3.10	< 0.05
	GUI+keyfr.	Keyfr. f+o	-0.14	3.36	< 0.05
	Keyfr. flick	Keyfr. f+o	0.01	0.63	< 0.05

less errors using either the *keyframe flick* (73% less) or *keyframe flick+overview* interface (91% less) than using the *GUI+keyframe* interface. Most errors using the *GUI+keyframe* interface were again of type E1 (see Table 6). Users also had the problem of dealing with the timeline at the top of the interface. In case of the *keyframe flick* interface, the few errors resulted from flicking too hesitantly. The same also holds for the *keyframe flick+overview* interface.

6.2.4 Navigation between Inter-related Videos

An overview on the amount of errors of the navigation between inter-related videos is given in Table 8. T-tests showed that the *2D flick* interface concept was significantly less error prone (81% less) than the *GUI+hyperlink* interface for both tasks ($p < 0.001$). Again, this is also due to the high usability of the gesture-based interface. The most common errors for the *GUI+hyperlink* interface were again of type E1 (see Table 8), due to the misplaced timeline. The errors with the *2d flick* interface were mostly slips.

Table 8. Amount of errors for the navigation in a collection of inter-related videos

	GUI+Hyperlink		2D Flick	
Task	1	2	1	2
E1	**42**	**26**	0	0
E2	19	9	**10**	0
E3	2	1	0	0
Slip	17	4	8	**5**
Sum	80	40	18	5

7. DESIGN IMPLICATIONS

Based on our analysis of different interface concepts within the design space, we derive implications for the design of mobile video browsers.

Support spatio-temporal browsing with flick interactions The evaluation showed that flick gestures are a highly efficient concept for temporal navigation within short video segments. Not only users navigated more quickly with this interface than with a time-slider or the tilt-based interaction concept. Moreover, the number of errors made with the flick interface was significantly lower than with the GUI-based interfaces. The evaluation also shows that by applying a spatial interaction metaphor in combination with simple, but highly efficient flick gestures, users are able to build up a mental model even of highly complex information spaces like inter-related digital libraries.

Support for discrete temporal navigation Although concepts for temporal navigation may afford continuous manipulation (e.g. continuously wiping over the display and therefore navigating through a video continuously), our observations have shown that users demand the possibility to navigate in discrete steps. By performing one gesture (e.g. flick or tilt) the video should be winded forth or back by a fixed amount of time, for instance 10 seconds. These gestures should be additive: By repeating the gesture several times, a larger amount of time can be navigated forth (or back respectively). Compared to a continuous interaction, where the video is being navigated as long as the gesture is performed, this has the advantage to offer better control and reversibility of the command.

Place GUI elements to be reachable by the user's thumb Classical GUI elements should be reachable by a user's thumbs. This guideline might appear straightforward. However, the evaluation shows that wide-spread interfaces (e.g. the iPhone movie player) do not follow this guideline. Our study shows that users most commonly hold the device in landscape mode for watching movies, since this offers the most screen real estate. Moreover, most of them utilized both hands to hold the device. In this case only the thumbs are able to interact with the interface. The rest of the hand is located *behind* the device. Consequently, the interaction is highly limited by the length of a user's thumb. If a horizontal timeline is used for navigation, such as in the iPhone video player, the timeline should be placed at the bottom of the interface. This allows reaching it with the thumbs and moreover prevents users from occluding the screen while using the timeline for navigation.

8. RELATED WORK

To the best of our knowledge, prior work has not contributed a detailed analysis of the design space of mobile video browsers. Although there has been a lot of research on interfaces for desktop computers, research on mobile interfaces can only hardly benefit from lessons learned in this design space. Desktop interfaces are typically GUI-based (e.g. [5,15,16]) and rely on traditional input modalities like a keyboard and a mouse. In contrast, mobile devices offer novel affordances like touch and tangible input and have significantly different form factors.

For *mobile* video browsing, Sun and Hürst [17] have developed various interfaces. Most notably, the ElasticSlider allows users to skim quickly through continuous video streams. This approach leverages a rubber band metaphor. By spanning the band, the playback speed is adjusted adaptively. This supports navigation at adjustable speeds during playback of a movie. However, this concept does neither enable selective interaction, nor the navigation between collections of inter-related videos.

Kamvar et al. developed the MiniMedia surfer [9], a mobile browser for small video segments. The browser supports keyword queries and users explore query results through key frames. The navigation completely relies on the designated keywords for each video segment. This is a major issue when trying to get an overview on a set of videos without knowing what to look for.

Dachselt and Buchholz [2] presented tilt-based interactions for mobile devices. While this was a source of inspiration for our *temporal tilt* interface, the mobile device is used as a remote control for media displayed on a distant screen. Moreover, the interaction with videos is not supported.

9. CONCLUSION

The aim of this paper was to adopt a broad view on the design space for mobile video browsing. We set up a design space that covers two dimensions: the broad interaction metaphor used in the interaction concept (GUI-based, gesture-based, physical) and the complexity of the navigation. This design space enabled us to systematically derive novel interaction concepts, both for efficient navigation within individual videos and for browsing collections of several inter-related videos.

For an in-depth evaluation of these interfaces, we conducted a controlled experiment with 44 participants and collected and analyzed more than 18 hours of video observations. Therefore, we were not only able to assess the usability of each interface, but also to identify where errors occur. The results provide empirical evidence that designers should leverage the novel capabilities of mobile devices, such as direct touch and inertial sensors. A more traditional GUI approach, as in this case the iPhone video player, is likely to lead to lower efficiency and is more error-prone. The usability error analysis shows that even a simple misplacement of interface elements can lead to the loss of internal locus of control and therefore to severe usability breakdowns. Moreover, the error analysis underlines the potential of gesture-based or physical interfaces for mobile video browsing. Our participants committed only little errors of type E1-E3 using either interface type. They mainly committed slips, if at all. These slips can be further reduced when users become more familiar with such novel interaction techniques.

Our analysis also provided the basis for design guidelines for mobile video browsers. By supporting spatio-temporal browsing metaphors and discrete temporal navigation and by placing interface elements carefully, designers can improve both usability and user experience of future mobile video browsers. As future work, we consider the further exploration of physical interaction techniques for mobile video browsing.

10. ACKNOWLEDGEMENTS

This work was funded by the German Research Foundation (DFG-GK-1223). The authors thank Attila Victor Achenbach, Roman Lissermann and Simon Olberding for their help with the experiment. The authors also thank the anonymous participants of the experiment. The authors are grateful to the anonymous reviewers for their valuable feedback.

11. REFERENCES

[1] Brooke, J. SUS: a "quick and dirty" usability scale. Usability Evaluation in Industry. London: Taylor and Francis, 1996.

[2] Dachselt, R. and Buchholz, R. Natural throw and tilt interaction between mobile phones and distant displays. In Proc. CHI '09. ACM, 3253-3258.

[3] Edwards, D.M. and Hardman, L. Lost in hyperspace: cognitive mapping and navigation in a hypertext environment, 90-105, Hypertext: theory into practice, 1999.

[4] Gong, J. and Tarasewich, P. Guidelines for handheld mobile device interface design. In Proc. DSI '04.

[5] Haubold, A., Dutta, P., and Kender, J. R. Evaluation of video browser features and user interaction with VAST MM. In Proc. Multimedia '08, ACM, 449-458.

[6] Hassenzahl, M., Platz, A., Burmester, M. and Lehner, K. Hedonic and ergonomic quality aspects determine a software's appeal. In Proc. CHI '00, ACM, 201-208.

[7] Hürst W., Welte M. and Jung S. An Evaluation of the Mobile Usage of E-Lecture Podcasts. In Proc. Mobility '07.

[8] iTunes U, http://www.apple.com/education/mobile-learning/

[9] Kamvar, M., Chiu, P., Wilcox, L., Casi, S., and Lertsithichai, S. MiniMedia surfer: browsing video segments on small displays. In Proc. CHI '04, ACM, 1371-1374.

[10] Karrer, T., Wittenhagen, M. and Borchers, J. PocketDRAGON: a direct manipulation video navigation interface for mobile devices. In Proc. mobileHCI '09, 1-3.

[11] Masson, M. E., Hill, W. C., Conner, J. and Guidon, R. Misconceived misconceptions? In Proc. CHI '88, 151-156.

[12] Norman, D. 2002. The Design of Everyday Things.

[13] O'Hara, K., Mitchell, A. S., and Vorbau, A. Consuming video on mobile devices. In Proc. CHI '07, ACM, 857-866.

[14] OpenCourseWare, http://ocw.mit.edu

[15] Rigamonti, M., Lalanne, D. and Ingold, R. FaericWorld: Browsing Multimedia Events Through Static Documents and Links. In Proc. INTERACT '07, Springer , 102-115.

[16] Shipman, F., Girgensohn, A., and Wilcox, L. Authoring, viewing, and generating hypervideo: An overview of Hyper-Hitchcock. ACM Trans. Multimedia Comput. Commun. Appl. 5, 2 (Nov. 2008), 1-19.

[17] Sun Q. and Hürst, W. Video Browsing on Handheld Devices-Interface Designs for the Next Generation of Mobile Video Players. In IEEE MultiMedia, 76-83, July-September, 2008.

[18] Truong, B. T. and Venkatesh, S. Video abstraction: A systematic review and classification. *ACM Trans. Multimedia Comput. Commun. Appl.* 3, 1 (Feb. 2007), 3.