

An Investigation of Textbook-Style Highlighting for Video

Matthew Fong* Gregor Miller† Xueqin Zhang‡ Ido Roll§ Christina Hendricks¶ Sidney Fels||

ABSTRACT

Video is used extensively as an instructional aid within educational contexts such as blended (flipped) courses, self-learning with MOOCs and informal learning through online tutorials. One challenge is providing mechanisms for students to manage their video collection and quickly review or search for content. We provided students with a number of video interface features to establish which they would find most useful for video courses. From this, we designed an interface which uses textbook-style highlighting on a video filmstrip and transcript, both presented adjacent to a video player. This interface was qualitatively evaluated to determine if highlighting works well for saving intervals, and what strategies students use when given both direct video highlighting and the text-based transcript interface. Our participants reported that highlighting is a useful addition to instructional video. The familiar interaction of highlighting text was preferred, with the filmstrip used for intervals with more visual stimuli.

Index Terms: H.5.2. [Information Interfaces and Presentation]: User Interfaces—; H.1.2. [Models and Principles]: User/Machine Systems—

1 INTRODUCTION

There is a long history of video being used as an educational tool, such as documentaries and visual tutorials. Streaming video sites, such as YouTube, have led to an enormous number of instructional videos across many topics, contributed from individuals on hobbies and interests up to formal education contributed from well-known universities. With the rise of online education frameworks (such as edX¹), video has become a core aspect of courses designed for self-learning. Massive Open Online Courses (MOOCs) use video substantially, often as the central teaching aid in place of lectures, to offer complete courses that are delivered over the web. In 2014, nearly 20,000,000 learners registered to at least one MOOC². Other sources, such as Khan academy³ offer more targeted videos for learners who seek information on specific topics.

One challenge with using large quantities of video to teach courses is providing students with effective management tools. Traditional playback controls (play, pause, seek) are generally used, which do not support recall, history or interval bookmarking. This requires the user to either note down timecodes for intervals or visually search the video when they view it again. However, video is a complicated medium to navigate, so it is unclear which types of interactions are helpful within a learning context. The first con-

tribution of this work is an investigation with university students to determine which type of video interfaces would help them with their learning experiences. We created a prototype interface with various features including filmstrip preview, view count visualization, personal viewing history, interval bookmarking and playlists. Students were individually interviewed while trying each feature and asked which they found to be useful or if they would modify or add aspects. Based on their feedback, we identified the most promising features and ran a focus group to obtain ideas on how to present them for instructional content.

Our second contribution is an investigation into highlighting of video using a textbook metaphor. Based on the outcomes of the interviews and focus group, we designed a prototype interface supporting video playback, view count visualization, transcript navigation and a 2D filmstrip. The transcript navigator and the filmstrip both allow “highlighting” of video intervals with various colours, exactly as one may do with a textbook. The goal of this prototype interface was to discover if the textbook metaphor of highlighting is effective and the use of highlights within instructional video is helpful for students.

There are various aspects of the video highlighting problem which we investigated with a qualitative evaluation:

- The main research question: Does the textbook metaphor of highlighting aid students in learning from instructional video?
- If so, then what type of content would they highlight?
- When do students highlight the content? (I.e. while viewing, in advance, or afterwards)
- Given a number of highlight colours, how do students assign meaning to these, and would they prefer pre-assigned meaning?
- How many colours should be available?
- Finally, what are the strategies used to highlight, or later find highlights? For example, did they use the transcript navigator more often for video with similar-looking visual content?

After a review of the literature we present the preliminary investigation, followed by a description of the new interface developed to test video highlighting and the evaluation performed to determine its utility.

2 RELATED WORK

The development of video interfaces within educational contexts is an important topic for contemporary learning. We first cover prior work from video interfaces for navigation and annotation, and then discuss pedagogical strategies and the use of video for learning.

2.1 Video Navigation

Current video navigation methods mostly use the familiar VCR-like controls (play, pause, seek, fast forward, rewind) and sometimes include chapter systems (e.g. DVDs) to skip to specific sections of a movie. Commercial video viewing systems have not progressed much further, despite significant research into new methods. Girgensohn et al. improved video thumbnails, and allowed users to directly manipulate the preview frame by moving a cursor along

* {email:mfong@ece.ubc.ca}

† {email:mfong@ece.ubc.ca}

‡ {email:xueqin@ece.ubc.ca}

§ {email:ido.roll@ubc.ca}

¶ {email:c.hendricks@ubc.ca}

|| {email:ssfels@ece.ubc.ca}

¹ <https://www.edx.org>

² <https://www.edsurge.com/news/2014-12-26-moocs-in-2014-breaking-down-the-numbers>

³ <https://www.khanacademy.org>

a timeline to find the desired clip [11]. Further work by Matjeka et al. [20] utilize multiple thumbnails to help users explore video. Kimber et al. [18], Dragovic et al. [8] and Karrer et al. [15] encouraged users to directly manipulate within-video content along its movement path to seek around video.

Moving towards traditional timelines, we see inherent problems as the length of video becomes larger. Representation of each time-step becomes too small to see and use. Work by Hürst et al. [13, 14] introduced the ZoomSlider, which hid parts of the timeline, and allowed users to shift the slider across the screen to seek across the video. The seeker bar is zoomable, and the granularity of seeking is dependent on the cursor's vertical position, allowing for higher accuracy in seeking for longer videos. This has been adopted commercially and can be seen in Apple's iOS video player.

There has been a recent trend towards allowing users to utilize their navigation histories in video, much like those found in web browsers. Al Hajri et al. visualize detailed video navigation history for general video in [3, 4], allowing users to view and explore previously seen video.

2.2 Video Annotation

There are several annotations tools being developed, some commercial, and some open source⁴, allowing users to mark up video, involving different methods of annotation, including text, ink, audio, or video. In research there has been work in video annotation tools designed for various devices, such as mobile [7], or stylus based [25, 27], collaborative annotation tools [26, 1, 9], as well as automatic annotation via computer vision techniques [12, 23]. The focus of the work in video annotation has been for general video. In education, annotation systems have been developed more for instructor use. To focus more on student roles in usage of video for education would be invaluable.

2.3 Video Education Strategies

A common use of video within education falls under the category of flipped classrooms, where students watch instructional videos at home to prepare themselves for class [5]. The videos are often created by the instructors, and include a standard "lecture", often using slides. The slides or lecture backdrop can lead to highly similar-looking content which can be difficult to navigate visually, so the use of a transcript for navigation was included in our interface.

While the use of video to support learning has become more popular, the style has remained largely intact as a linear lecture [28]. This can be improved in some simple ways; for example, people learn better from instruction that includes text *and* video, compared to either in isolation [6]. Additional evidence suggests that learning is improved when students have control over their learning and can watch videos in segments, at their own pace [6]. Our viewer is designed to accommodate this, and provide students with the ability to segment the video themselves for later review.

To our knowledge, highlighting, a major aspect of textbook learning, has not yet been tested on video. Learners learn better when they can highlight information that they find valuable [19], which we attempt to provide them in our interface.

2.4 Video Interfaces for Education

Mertens et al. developed the idea of video footprints, a detailed view count for the length of the video based on personal viewing, visualized as a heatmap timeline [21]. This was presented together with temporal bookmarks (at a single time instant) and navigation using either slides or time. These ideas were further explored by Kim et al. [17], through the use of interaction data (including video footprints) to help students navigate instructional video. Frequently-viewed regions are isolated within a histogram,

⁴<http://www.annotations.harvard.edu/icb/icb.do?pageid=icb.page466612>



Figure 1: The Video Library screen. Each video has a title, an author, and a description. The second video "Sorting - Ep 05 - Insertion Sort" currently being edited.



Figure 2: The Filmstrip as seen in the Video Library when the cursor is placed over a video. This at-a-glance view shows the user what parts of the video has been highlighted (top, multicolour) and what parts they have seen (bottom, red/orange). The user can also move the mouse across the Filmstrip to preview the video.

and video summaries can be generated based on personal or crowd-sourced interactions. The interface also supports transcript visualization and temporal bookmarking. Our work extends these interfaces with interval bookmarking (highlighting) coupled to a highlighted transcript, together with a heat-map based view count to aid navigation. Pavel et al. presented an interface to create video digests, smaller chunks of video with previews and text summaries, to help people skim and browse long videos [24]. The interface is based on a transcript of the video, with automatic and crowd-sourced first passes and a manual refinement step if needed. The creation of chapters using transcripts is similar in principle to our approach, using video clips as the basis for subsequent navigation.

RIMES explored the use of interactive multimedia exercises embedded within lecture videos, with student feedback recorded using video, audio and sketching for the instructor to review later [16]. Each video was self-contained with its exercises, with annotation available to the student to answer questions, but no mechanisms were present for later referral. Monserrat et al. presented an education environment called L.IVE, which provided students with the ability to discuss parts of the video with temporal comments (marked at a single time in the video), along with assessments included with the video [22]. This provided students with an integrated viewing and assessment environment, however the navigation was based on the traditional controls and did not organize the comments to aid navigation.

3 PRELIMINARY INVESTIGATION

As a start to determining the usefulness of current video interfaces we decided to perform interviews and focus group studies on students who would be able to provide proper feedback. We chose interfaces that we believed would support users in studying: the navigation history visualizations in [4], the revisitation heatmaps from [17], a filmstrip type visualization from [20], and a form of authoring functionality from [10]. In the following studies, we pre-

sented a combination of these interfaces to students and collected feedback to inform the design of a video highlighting solution.

3.1 Interviews

We conducted interviews with 12 participants to identify how students watched videos for their learning online, and in regular classes. The participants were aged 19 to 33, three male, eight female, and all the participants had taken at least one course with video. We introduced to the participants the idea of the flipped classroom, where video lectures are watched at home, and “homework” is completed in class, and used that as the context in which the participants should think about answering the questions we posed to them. We first looked at the methods that the participants used to study from videos. Following that, we demonstrated implementations of interfaces (navigation history [4], revisitation heatmap [17], filmstrip [20], authoring [10]) to the participants and asked them about their thoughts on how they would use them while studying, and whether they would be improvements upon their current video learning experiences.

We found that studying using video was essentially difficult, and the current tools provided (which were most often YouTube, Khan Academy, and Coursera), were not sufficient in allowing students to study well from the video. For example, students needed to manually search through a linear timeline to find portions of video that they were interested in as they could not save them. Other students were inclined to write down time stamps of portions of video that they did not understand and manually seek to them later. Search in the video was very cumbersome and the lack of being able to visualize the video served as a detriment to their studying. Some students found themselves watching the videos, taking notes, and not watching the video again.

Out of the four different interface elements that were shown to the participants, the most liked element was the revisitation heatmaps, followed by the playlists, the history and then finally the filmstrip. It was stated that being able to see which parts of the video needed more attention would be beneficial to their reviewing process. Furthermore, they stated that being able to see others’ heatmaps on top of their own would be beneficial. The playlist feature was useful in that instead of having to write down timestamps and search through the video for a specific part, they would be able to save the video they want and create their own review videos. The history proved to be a little confusing and with the limited demonstration, the participants were not given enough time to familiarize themselves with the history.

3.2 Focus Group

We conducted a focus group study amongst a group of students who had just completed an offline philosophy class that was taught partially with video using a standard video player (YouTube). In this study, we intended to allow the participants to use the interface with videos that they already studied from. This would allow them to compare YouTube and the experimental interfaces with the same videos they have already watched. We had seven participants in this focus group aged 18 to 26, two male, and five female. This time, the participants were given the interfaces to use on their own laptops and they were given a tutorial for each interface and given time to use it and acclimatize to the interface. While the participants used the interface, we asked them to speak aloud any frustrations or ideas they had, to better inform us of a design of an interface that can support their studying needs.

The focus group showed that being able to view a history of a user’s behaviour ([4]) was useful, however, the interface was difficult to use and the students were frequently confused. They found that because the videos are educational and are essentially slideshows, the thumbnails being generated were too visually similar and useless in differentiating between different video portions

of video. In fact, the same problem revealed itself in the case of the filmstrip ([20]), since the same frame can be seen across lengths of time resulting in a filmstrip that, while interactive, did not provide visual feedback of its interaction.

However, being able to select intervals of video for manipulation ([10]) was useful in terms of organization. The idea of being able to keep track parts of the video that have been seen more often, or parts of the video that have not been seen at all ([17]) was intriguing to the participants of the focus group. They wanted more control over it, and the ability to emphasize or de-emphasize certain parts of video without having to watch it over and over. Furthermore, they liked the idea of the playlist and being able to use it as an organizational structure mechanism.

Looking at the implications of user-adjustable emphasis and video organization tools, we were led to the development of the proposed textbook metaphor for highlighting video and the prototype interface described in the following section.

4 INTERFACE

The main idea behind the interface is to allow users to manipulate emphasis on certain parts of video and be able to use that to organize their video in their own way. Thus, the interface was designed to allow users to highlight video, and then play back those highlights. In this way, they are emphasizing video (by highlighting it), and creating playlists (by being able to play back highlighted parts). We designed a video player around this concept and took cues from physical highlighting tools as well as digital ones found in document readers. The interface consists of two screens and three major elements. The first screen, shown in Figure 1 allows the user to explore the videos offered by the interface. Clicking on one of the videos brings the user to the video player screen, shown in Figure 3. This screen consists of the three elements: the filmstrip (red), the player (green), and the subtitle viewer (blue). This section will describe the design guidelines and each of the interface elements in detail.

4.1 Video Library

The video library shows a list of videos offered by the interface. The user can see at a glance, the title, the author, the description, and a preview thumbnail. Upon putting the cursor over an item, a filmstrip, described in the next section, appears, allowing the user to quickly explore the visual contents of the video, see the highlighted parts of the video, as well as see which parts of the video have been seen. This can be seen in Figure 2. Each video’s details can also be edited to the user’s liking.

4.1.1 Filmstrip

The filmstrip is a set of thumbnails from the video arranged side by side, each representing a portion of video. As the width of the filmstrip represents the entire length of the video, each n thumbnail represents $1/n$ of the video. In the case of the video library, there are three thumbnails, and each thumbnail represents one-third of the video. Moving the cursor over top causes a floating timestamp to appear, and changes the corresponding thumbnail to show the frame represented by the horizontal location of the cursor. The initial image is the first frame of the represented interval.

On the top side of the filmstrip is a bar indicating the highlights made in the video across the filmstrip, with respect to the ratio between the length of the video and the width of the filmstrip. In pilot tests, users found that in certain videos, slide-based ones in particular, the filmstrip resembled a set of slides instead of an actual timeline. In order to rectify this, we chose to insert time-ticks underneath at the bottom, distributing them evenly to a minimum of either one tick per second, or one tick every 10 pixels across the width of the filmstrip. Along the time ticks, there are two timestamps placed at $1/3$ and $2/3$ of the width of the filmstrip to indicate

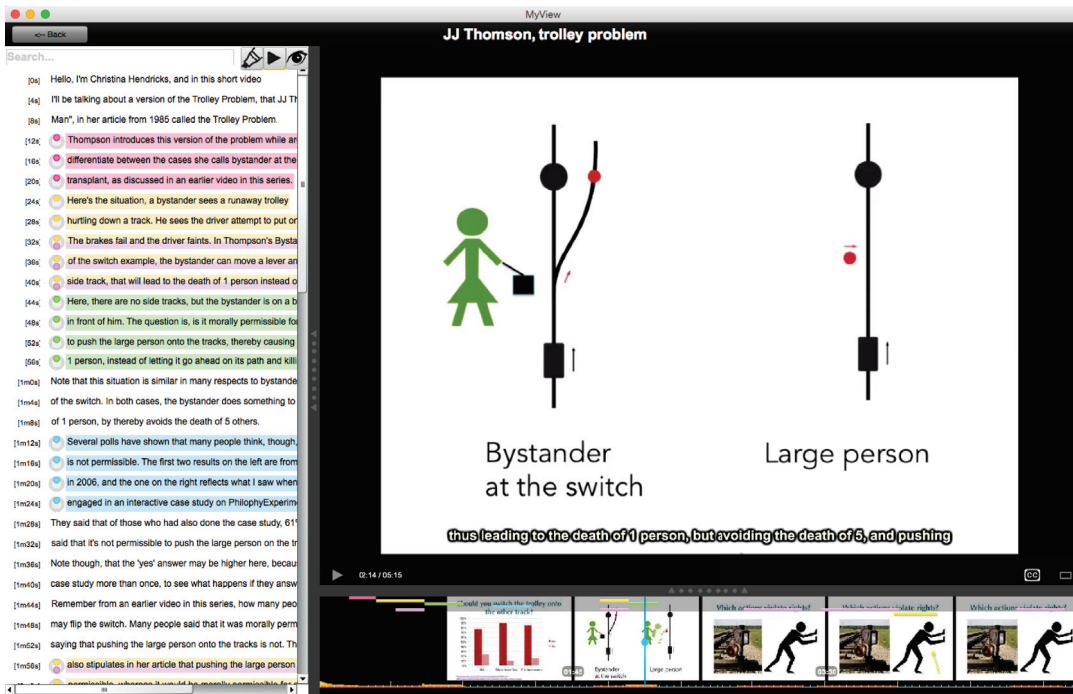


Figure 3: The main video player view. The filmstrip (red), the player (green), and the subtitle viewer (blue), reside here to help the user watch, review, and highlight the video.

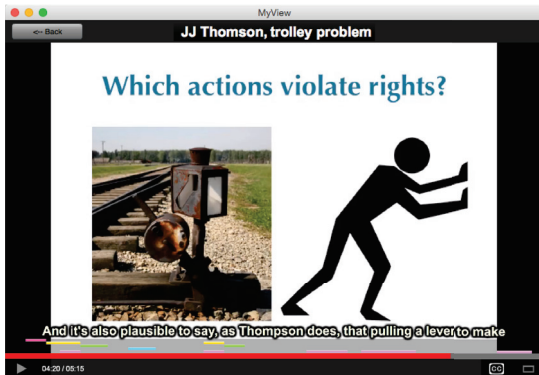


Figure 4: The main player in fullscreen view. In this mode, all the other interface elements are hidden, reducing distractions from the video. The highlights made in the video are visible at the bottom.

time. Underneath the time ticks, the filmstrip shows a heatmap of the user's visitation of the video. The height of the graph is logarithmic, but never overlaps with the bottom of the thumbnail, and the colour ranges from red to orange to yellow, with red being least seen and yellow being most seen. This visualization is similar to [17], and deals with scalability better, because the usage of logarithmic heights ensures that portions of video with low view count are still visually represented. It is also more precise than [2] because the horizontal axis of the graph is linear and the locations of the thumbnail are stationary. The resolution of visitation is stored on a per-second basis. When that second in the video has been fully watched, the application adds to the view-count for that second. This resolution was a good balance between performance and

visual representation on the screen.

4.2 Video Player

The video player, Figure 3, allows the user to view and markup the video. It consists of three major elements, the player, the filmstrip, and the subtitle viewer.

4.2.1 Player

The player, shown green in Figure 3 is the main focus of the interface, and is typical of those usually found in a video player. On the bottom is a toolbar that houses video controls, allowing the user to play (▶) or pause (⏸) the video, view the currently playing time, turn on and off closed captioning (CC), or make the video fullscreen (⏏). When the video is made fullscreen, a seekbar appears on top of the toolbar, allowing the user to seek without view of the Filmstrip, which will be described later on. Furthermore, when the player is made fullscreen, a bar of highlights appears on the top above the seek bar. This can be seen in Figure 4.

4.2.2 Filmstrip

The filmstrip in the video player is very similar to the filmstrip found in the video library, with some extended functionality. As we are in the player, the filmstrip provides the user with the ability to highlight intervals. By clicking and dragging across the thumbnails, the user can select intervals, and then a button will pop up. When the cursor is over the button, more buttons will pop up, allowing the user to select the colour they want to highlight in.

The filmstrip in the player can also be expanded. By clicking dragging on the bar above it, the filmstrip can expand into multiple rows, and the scale of time is expanded across all the filmstrips. For example, in Figure 6, there are three filmstrips and the width of each filmstrip represents one-third of the video. Some of the filmstrips ends and beginnings are jagged, indicating that they represent only parts of the video and not all of it.

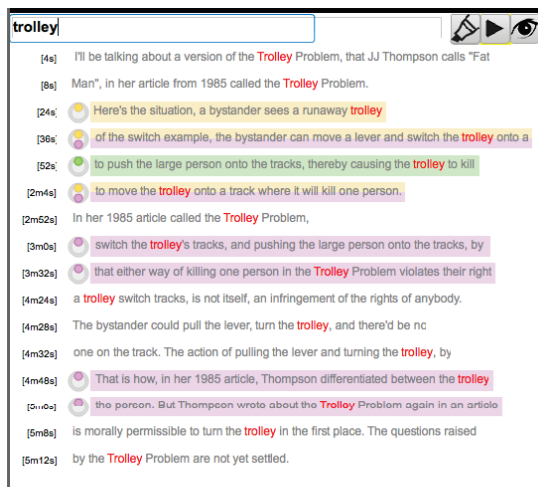





Figure 5: Subtitle Viewer Toolbox expanded to show the options of colours available to highlight.

4.2.3 Subtitle Viewer

The subtitle viewer, shown on the left in Figure 3 in blue, is a transcript of everything said in the video. This provides the user with an overview of video in textual form, allowing them to quickly scan through the spoken content of the video. As the video plays through, the text being currently spoken turns red and acts like a playhead. On the left of each caption is a timestamp, which the user can click, causing the player to seek the the time indicated.

Like the filmstrip, the user can click and drag across text to select it, and using the  button, the user can highlight the selected text in the chosen colour. Like in the filmstrip, text can be highlighted in more than colour, and each colour will be stacked on top of each other underneath the text. Once a line of text has been highlighted, a miniature widget appears next to the text indicating to the user the colour the text has been highlighted. This is useful when the text has been highlighted in multiple colours, as seeing which colour underneath the text can be difficult.

The  button allows the user to playback highlights and the  hides the highlights within the text. Included is also search functionality, allowing the user to search for text within the subtitles. When the user enters search terms, captions that do not contain any matches disappear and captions with a match remain on the screen. The text is greyed out with the exception of the matching term within the text, which is marked in red. The result of a search can be seen in Figure 5.

4.3 Design Strategies

In an effort to maintain consistency across the entire interface, there were some rules that we adhered to when designing the interface. For each of the interface elements in the video player, selections and highlights performed in one element is reflected in the others instantly. For example, selecting an interval in the subtitle viewer causes the corresponding interval in the filmstrip to be selected as well. This provides consistency and visual feedback for the user in more than one location. Performing a highlight causes the same effect on both interface elements as well.

The colour of the highlights on the filmstrip and subtitle viewer are always in order. We experimented with having colours appear in the order that they were created but found in pilot tests that having the order change across different videos to be confusing, so we opted to make the colours consistently go from red to purple.

The height of the thumbnails in the filmstrip as set to a minimum of 100 pixels, and the width was adjusted to the aspect ratio of the video itself. In pilot studies, we found that it was difficult to discern details in thumbnails smaller than 100 pixels. Furthermore, thumbnails were stripped of any black bars surrounding the thumbnail, ensuring that space on the filmstrip was used as efficiently as possible to convey the contents of the video.

5 EVALUATION

We ran a qualitative user study to investigate the usability and utility of the two interface elements for highlighting video presented in our system. We observed user reactions to the system through a questionnaire, as well as observed and analyzed various usage patterns that came out of having users watch videos using the interface.

5.1 Participants

11 volunteers participated in the experiment: 6 male and 5 female, ranging in age from 18 to 33. They were monetarily compensated for their time. All participants have taken at least one course online, and all but one have taken at least one course that used video as a teaching medium.

5.2 Design and Procedure

Each participant was exposed to the three methods for highlighting described above. The participants were given eight different videos to choose from, where each video was 2 to 3 minutes long and of the educational category. The videos were found on YouTube:

1. How chronic obstructive pulmonary disease (COPD) develops⁵
2. How to Balance Chemical Equations⁶
3. How To Read Music - Two Minute Music Theory⁷
4. Learn Japanese Verb Groups⁸
5. Schrödinger's Cat⁹
6. Sorting - Ep 05 - Insertion Sort¹⁰
7. Supply and Demand¹¹
8. The Trolley Problem (with a Twist)¹²

By being able to select a video of their choosing, we hoped they would choose a video that is of interest to them and encourage them to pay attention to the video and use the interface in a way that would increase their enjoyment of the interface. They were asked to watch the video and pretend they were studying the video for a course and to highlight portions of video as needed. The experiment took approximately 30 minutes.

The evaluation began by introducing the participants to the idea of being able to markup video as they would in a textbook, and a description of the experiment. We first started by loading up a demonstration video, and introduced the *Subtitle Viewer* and its highlighting and search functionality. The Filmstrip was disabled during this section of the experiment. The participant was then asked to pick a video of their choice, watch it as if they were studying it, and highlight it as necessary with the Subtitle Viewer. Once finished the

⁵<http://www.youtube.com/watch?v=iYjKZHmzWEA>

⁶<http://www.youtube.com/watch?v=8KH3laR2iR4>

⁷<http://www.youtube.com/watch?v=GiM2OKtACAQ>

⁸<http://www.youtube.com/watch?v=nOXuIYVzyL4>

⁹<http://www.youtube.com/watch?v=IOYyCHGWJq4>

¹⁰<http://www.youtube.com/watch?v=3orUYqcaEEQ>

¹¹<http://www.youtube.com/watch?v=WdPI3hKUJYo>

¹²<http://www.youtube.com/watch?v=kKHOPw6tpd4>

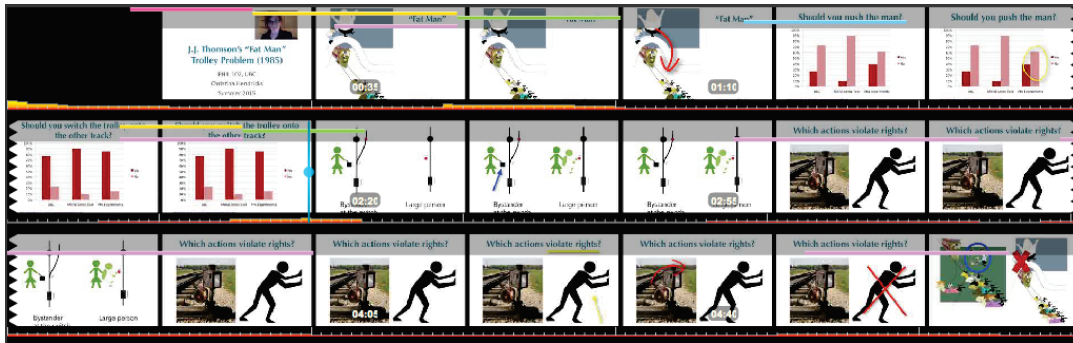


Figure 6: The Filmstrip, when resized, will split into multiple rows, each representing a portion of the video. Here, each row represents one-third of the video. Highlights are shown on the top of each row, and revisitation heatmaps are shown on the bottom of each row.

Question	Score
Easy to use	4.18
Powerful	4.27
Flexible	4.09
Aesthetically pleasing	3.82

Table 1: General reactions to the system, Likert scale from 1 to 5.

experimenter asked the participants three questions about the video, that they were told to find answers to in the video. The questions were content-based questions and the answers were located around one-third, one-half, and two-thirds of each video. The process was repeated for the Filmstrip, each time with a different video, with the opposing interface elements disabled. After introducing and using three elements, the participants were asked to watch one final video with the entire interface enabled. They were then asked to complete a questionnaire detailing their experience with the interface. The questionnaire contained 10 questions that were answered on a 5-point Likert scale. The experimenter then conducted a post-experiment interview asking about their experience with highlighting video, as well as their highlighting experiences in ebooks and physical textbooks.

6 RESULTS AND DISCUSSION

The questionnaire consisted of four questions which asked the participants to rate the overall reactions of the system (Table 1), and three questions for the Subtitle Viewer and the Filmstrip (Table 2). These are modified QUIS questions. Results derived from the questionnaire showed that the system was easy to use ($M = 4.18$), and learning to use each of the interface elements was easy (Subtitle Viewer $M = 1.18$, Filmstrip $M = 1.36$). General reactions to the video can be seen in Table 1. Questions related to each particular interface element are shown in Table 2. T-tests were conducted on each question. For “Learning was difficult”, there was no significant difference ($t = 0.452$); for “Distracting”, there was a significant difference ($t = 0.045$); for “Utility of highlighting”, there was no significant difference ($t = 0.126$). We also asked users to rank their preference of interface element to perform the highlighting and found a statistically significant difference ($\chi^2(1) = 11.000, p = 0.001$), where the Subtitle Viewer was the preferred method.

The Subtitle Viewer was found to be preferred over the Filmstrip for highlighting, and also found to be less distracting while watching a video. Furthermore, while participants found both easy to use and both had useful highlighting functions, the Filmstrip was significantly more distracting than the Subtitle Viewer. The participants who thought the Filmstrip was distracting cited the large playhead moving across as the video played as well as the thumb-

Question	Subtitle Viewer	Filmstrip
Learning was difficult	1.18	1.36
Distracting	1.27	2.18
Utility of highlighting	4.27	3.45

Table 2: Reactions to interface elements, Likert Scale from 1 to 5.

nails being distracting. The participants who thought the Subtitle Viewer was distracting cited the large amount of text on the screen. When shown that it was possible to hide the panel, they were satisfied. In most cases however, participants would use the Subtitle Viewer when asked to search for something in the video.

6.1 Highlighting

Participants utilized many different strategies of highlighting among the different interface elements. When given only the Subtitle Viewer, the participants utilized the following techniques:

1. Pause the video and highlight the past video. (1 participant)
2. Watch entire video first, then go back and highlight. (4 participants)
3. While the video is still playing, highlight parts that have already been played. (7 participants)
4. Read ahead and highlight ahead of the still playing video. (1 participant)

The first three techniques are all retroactive highlighting, but having the Subtitle Viewer allows users to look ahead of what is currently playing and evaluate the importance of it. For users, this means that going through a video can be done much faster than real-time, allowing them to be more efficient in their studying. Some participants in our preliminary interviews stated that when they watch online videos for studying, they often speed up the video 1.5 to 2 times faster. This current system does not have this, but allowing users to read the lecture at their own pace allows them to control the speed of their learning. In fact, a couple participants stated that they read ahead of the video, and one participant even skipped forward in the video, having evaluated that a portion of video was not worth watching. The majority of the participants chose to watch the video, and evaluated each sections importance as they watched it, highlighting past sections of video as they thought necessary.

When given only the Filmstrip, participants restricted themselves to retroactive highlighting, such as:

1. Pause the video and highlight the past video. (1 participant)

2. Watch the entire video first, expand the Filmstrip, then go back and highlight. (7 participants)
3. While the video is still playing, highlight parts that have already been played. (3 participants)
4. Rewind and highlight the appropriate part while the video is playing. (1 participant)

While the participants remembered to highlight using the Subtitle Viewer, they had to be prompted to highlight when they were given the Filmstrip. When given both interfaces participants, only two participants decided to use the Filmstrip at all to highlight, after having watched the entire video and going through again to highlight important points. Evidently, the highlighting task is not only encouraged by the Subtitle Viewer, but easier and more convenient to do as well. When asked about highlighting in the Filmstrip, participants said that it was difficult to see what was being highlighted due to the lack of knowledge about what is being said during the highlighted interval. Many participants needed to be reminded to highlight the video using the Filmstrip (as part of the task), and consequently, it can be seen that most participants watched the video and then went back to highlight the video. This is reflected in the lower “utility of highlighting” score for the Filmstrip in the questionnaire results when compared to the Subtitle Viewer.

6.2 Search

The search task required the participants to find portions of video using either interface element. Participants had more trouble with the search task using the Filmstrip than the Subtitle Viewer. This is due to the amount of distinguishable information that can be displayed by each interface element. In the Subtitle Viewer, the text is easily distinguishable, and it is easy to pick out certain words, especially when aided by the search function. The Filmstrip, while proved in other works to be useful for general video which contains many scene changes, is difficult to use due to the nature of educational videos. However, in the case that the video contained screens of varying scenes, the searching using the Filmstrip became far easier. In the experiment, sometimes, depending on how the Filmstrip was expanded, the exact frame containing the answer to the search task appeared in the Filmstrip and participants were able to quickly hone in on the correct location.

As far as the search task itself, when the participants were using the Subtitle Viewer, they:

1. Watched through the video again. (1 participant)
2. Skimmed through text. (6 participants)
3. Used the text search. (1 participant)
4. Looked for the highlight when they thought it was highlighted. (4 participants)

Looking at the methods used, one might assume that using the text search is not only the easiest option, but also the most used option. However, participants were more likely to scroll to the portion of the text they thought the answer was in, and start skimming the text, searching for keywords manually. In these cases, participants stated that because they knew approximately where the answer was, it was easier to skim the text than perform a text search. This is because the text search required the participant to think of an appropriate keyword to search, and because the non-matching results were removed, breaking the linear flow of the text, and it was not clear where the resulting search results resided in relation to the length of the entire video. However, when the participant knew that the question that was asked had been previously highlighted, they searched for a highlight and were generally able to find it instantaneously.

When participants were using the Filmstrip, they:

1. Expanded the Filmstrip and searched for visual information in the thumbnails. (9 participants)
2. Looked for highlights if they knew they highlighted the part. (5 participants)

Finally, when given both the Filmstrip and the Subtitle Viewer, participants did the following:

1. Searched to approximate point in the Filmstrip and played the video. (3 participants)
2. If the portion was highlighted, used the highlight playback function. (5 participants)
3. Used the text search. (3 participants)
4. Looked for the highlights in the Subtitle Viewer. (5 participants)

In terms of highlighting strategy, participants stated that they would highlight what they didn’t understand, what they thought were key points in the video, as well as review sections in the video. Oftentimes, when the participants were asked questions about something that they highlighted, they recognized it and were able to find the location of the video they were looking for, whether it was in the Filmstrip or the Subtitle Viewer. This indicates that the act of highlighting video acts as a memory aid for recall.

When asked about highlight colours, participants had varied answers, the most popular of which, was that choosing colours was random. In this case, there was no pattern in which they used highlight colours. For some, they would highlight things in the order that they were given. For example, the first thing they would highlight would be red, the second highlight would be in yellow, the third highlight would be in green, etc. It is worth noting that of the participants who utilized this method of highlighting were the quickest to find the answers to the experimenter’s questions, immediately know which colour to look for and could very quickly find the appropriate time. One participant stated that they would use red for important points in the video and the rest of the colours would be random. Two participants stated that they would colour code their highlights based on the topic within the video, for example, definitions, or instructions sets. All participants except for one stated that the five colours included within the system was adequate for their needs. The participant who wanted more stated that the number of colours should be an even number, for example, four, six, or eight colours. The reasoning behind this was that she wanted contrasting colours so that she could highlight things positively or negatively. When asked if they highlight digital ebooks or PDFs, nine participants said they did; the other two said that they did not.

The textbook metaphor and the use of highlighting was the more popular option utilized when participants were asked to search for video clips, whether it was manually looking through the highlights in the Filmstrip or the Subtitle Viewer, or by using the highlight playback function. The high usage of the highlight playback function for search indicates its usefulness in that context. Participants asked for the ability to skip to the next highlight while using this feature. Finally, it should be noted that for educational video at least, when presented with both the Filmstrip and the Subtitle Viewer, search was performed more often using text possibly due to the lack of movement and scene changes in educational video.

7 CONCLUSIONS AND FUTURE WORK

We presented a novel interface for highlighting video in both text and video representations. We performed two preliminary studies to inform us of what kind of interfaces work in a learning and education context, and developed an interface that allowed users

more control over organizing the video they watched. This interface adopts the textbook metaphor, using highlighting and text as its main focus, which was made to provide users with familiarity, making it easier to learn. We then evaluated the interface and compared highlighting and search functions across both the Filmstrip and the Subtitle Viewer and found that the Subtitle Viewer was more preferred, both in a questionnaire, and in practice. In the context of educational videos, we can conclude that videos that include transcripts offer more utility for users and will allow them to highlight, search, and review the video more easily.

In the future, we intend to investigate collaborative video interfaces that can be used in and out of classrooms. This would allow students to share information and help each other, as well as allow the instructor more engagement with students. Many of the participants in the preliminary interviews stated that the face-to-face contact was invaluable, and we believe that adding interaction between instructors and students in a video interface will aid in that respect. We also intend to test the interface on a larger scale, and deploy it a real world setting at a college or university. That will allow for even more feedback on the emerging video learning interfaces.

REFERENCES

- [1] G. D. Abowd, M. Gauger, and A. Lachenmann. The family video archive: An annotation and browsing environment for home movies. In *Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*, MIR '03, pages 1–8, New York, NY, USA, 2003. ACM.
- [2] A. Al-Hajri, M. Fong, G. Miller, and S. Fels. Fast forward with your vcr: Visualizing single-video viewing statistics for navigation and sharing. In *Proceedings of Graphics Interface 2014*, GI '14, pages 123–128, Toronto, Ont., Canada, Canada, 2014. Canadian Information Processing Society.
- [3] A. Al-Hajri, G. Miller, S. Fels, and M. Fong. Video navigation with a personal viewing history. In *Human-Computer Interaction – INTERACT 2013*, volume 8119 of *Lecture Notes in Computer Science*, pages 352–369. Springer Berlin Heidelberg, 2013.
- [4] A. Al-Hajri, G. Miller, M. Fong, and S. S. Fels. Visualization of personal history for video navigation. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 1187–1196, New York, NY, USA, 2014. ACM.
- [5] J. L. Bishop and M. A. Verleger. The flipped classroom: A survey of the research. In *ASEE National Conference Proceedings*, 2013.
- [6] R. C. Clark and R. E. Mayer. *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley and Sons, 2011.
- [7] B. C. Cunha, O. J. Machado Neto, and M. d. G. Pimentel. Movia: A mobile video annotation tool. In *Proceedings of the 2013 ACM Symposium on Document Engineering*, DocEng '13, pages 219–222, New York, NY, USA, 2013. ACM.
- [8] P. Dragicevic, G. Ramos, J. Bibliowicz, D. Nowrouzezahrai, R. Balakrishnan, and K. Singh. Video browsing by direct manipulation. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 237–246, New York, NY, USA, 2008. ACM.
- [9] R. Fagá, Jr., V. G. Motti, R. G. Cattelan, C. A. C. Teixeira, and M. d. G. C. Pimentel. A social approach to authoring media annotations. In *Proceedings of the 10th ACM Symposium on Document Engineering*, DocEng '10, pages 17–26, New York, NY, USA, 2010. ACM.
- [10] M. Fong, A. Al Hajri, G. Miller, and S. Fels. Casual authoring using a video navigation history. In *Proceedings of Graphics Interface 2014*, GI '14, pages 109–114, Toronto, Ont., Canada, Canada, 2014. Canadian Information Processing Society.
- [11] A. Girgensohn, J. Boreczky, and L. Wilcox. Keyframe-based user interfaces for digital video. *Computer*, 34(9):61–67, Sept. 2001.
- [12] D. B. Goldman, C. Gonterman, B. Curless, D. Salesin, and S. M. Seitz. Video object annotation, navigation, and composition. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, UIST '08, pages 3–12, New York, NY, USA, 2008. ACM.
- [13] W. Hurst, G. Gotz, and T. Lauer. New methods for visual information seeking through video browsing. In *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, pages 450–455, 2004.
- [14] W. Hurst and P. Jarvers. Interactive, dynamic video browsing with the zoomslider interface. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 4 pp.–, 2005.
- [15] T. Karrer, M. Weiss, E. Lee, and J. Borchers. Dragon: a direct manipulation interface for frame-accurate in-scene video navigation. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 247–250, New York, NY, USA, 2008. ACM.
- [16] J. Kim, E. L. Glassman, A. Monroy-Hernández, and M. R. Morris. Rimes: Embedding interactive multimedia exercises in lecture videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1535–1544, New York, NY, USA, 2015. ACM.
- [17] J. Kim, P. J. Guo, C. J. Cai, S.-W. D. Li, K. Z. Gajos, and R. C. Miller. Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 563–572, New York, NY, USA, 2014. ACM.
- [18] D. Kimber, T. Dunnigan, A. Girgensohn, F. Shipman, T. Turner, and T. Yang. Trailblazing: Video playback control by direct object manipulation. In *International Conference on Multimedia and Expo*, pages 1015–1018. IEEE, July 2007.
- [19] D. Leutner, C. Leopold, and V. den Elzen-Rump. Self-regulated learning with a text-highlighting strategy. 215(3):174–182, 2007.
- [20] J. Matejka, T. Grossman, and G. Fitzmaurice. Swifter: Improved on-line video scrubbing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1159–1168, New York, NY, USA, 2013. ACM.
- [21] R. Mertens, H. Schneider, O. Müller, and O. Vornberger. Hypermedia navigation concepts for lecture recordings. In *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, volume 2004, pages 2840–2847, 2004.
- [22] T.-J. K. P. Monserrat, S. Zhao, Y. Li, and X. Cao. Live: An integrated interactive video-based learning environment. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 185–186, New York, NY, USA, 2014. ACM.
- [23] K. Ntalianis, A. Doulamis, N. Tsapatsoulis, and N. Doulamis. Human action analysis, annotation and modeling in video streams based on implicit user interaction. In *Proceedings of the 1st ACM Workshop on Analysis and Retrieval of Events/Actions and Workflows in Video Streams*, AREA '08, pages 65–72, New York, NY, USA, 2008. ACM.
- [24] A. Pavel, C. Reed, B. Hartmann, and M. Agrawala. Video digests: A browsable, skimmable format for informational lecture videos. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 573–582, New York, NY, USA, 2014. ACM.
- [25] G. Ramos and R. Balakrishnan. Fluid interaction techniques for the control and annotation of digital video. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, UIST '03, pages 105–114, New York, NY, USA, 2003. ACM.
- [26] M. Riegler, M. Lux, V. Charvillat, A. Carlier, R. Vliegndhart, and M. Larson. Videojot: A multifunctional video annotation tool. In *Proceedings of International Conference on Multimedia Retrieval*, ICMR '14, pages 534:534–534:537, New York, NY, USA, 2014. ACM.
- [27] J. a. Silva, D. Cabral, C. Fernandes, and N. Correia. Real-time annotation of video objects on tablet computers. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, pages 19:1–19:9, New York, NY, USA, 2012. ACM.
- [28] A. M. F. Yousef, M. A. Chatti, U. Schroeder, M. Wosnitza, and H. Jakobs. Moocs - a review of the state-of-the-art. In *Proceedings of the Computer Supported Education Conference*, volume 3, pages 9–20, 2014.