Video Annotations in Helping Locate In-Video Information for Revisitation

by

Min Li

B. Eng, University of Science and Technology Beijing, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Electrical and Computer Engineering)

The University of British Columbia
(Vancouver)

June 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Video Annotations in Helping Locate In-Video Information for Revisitation

submitted by **Min Li** in partial fulfillment of the requirements for the degree of **MASTER OF APPLIED SCIENCE** in **Electrical and Computer Engineering**.

Examining Committee:

Dr. Sidney Fels, Electrical and Computer Engineering
Supervisor

Dr. Dongwook Yoon, Computer Science
Supervisory Committee Member

Dr. Konstantin Besnozov, Electrical and Computer Engineering
Supervisory Committee Member

Abstract

Rewatching video segments is common in video-based learning, and video segments of interest need to be located first for this rewatching. However, learners are not well supported in the process of locating in-video information. To fill this gap, the presented work explores whether video annotations are effective in helping learners locate previously seen in-video information. A novel interface design consisting of two components for learning with videos is proposed and tested in the task of locating in-video information: an annotating mechanism based on an integration of text with video, and an annotation manager which enables the learner to see all annotations he/she has made on a video and provides quick access to video segments. A controlled lab experiment with 16 undergraduate students as subjects was carried out. Experiment results suggested that the use of video annotations significantly reduced time spent on searching for previously seen video segments by about 5 seconds ($p < 0.05$), and subjects spontaneously used the proposed annotation manager 3 times more often in the information-seeking process than the traditional method of finding video segments by sliding through the video timeline (9:3). Qualitative data from surveys and interviews indicated that both the annotation process and annotations on the proposed interface were perceived as helpful for learning with videos. Thus, video annotations in the proposed interface are effective in reducing search time for in-video information; the annotation process, annotations created by the learner, and the proposed annotation manager play important roles in the information-seeking process.

Lay Summary

With online video learning being more and more popular among students, tools to facilitate such learning experiences are not keeping pace. When learning by reading a book, there are many devices available to improve the learning experience: highlighters, pens for note-taking, bookmarks, tables of contents, and indexes. However, when learning by watching a video, the learner is restricted aside from clicking on the video timeline and pressing a very limited number of buttons provided by the video player. Rewatching a video is common, especially in a learning context similar to rereading textbooks. However, there are also very few features designed to support learners in this rewatching process.

Thus, a novel interface of video learning with richer features is proposed and tested. The interface has been found to be highly preferred by learners and effective in helping learners find previously seen video segments to rewatch.

Preface

The novel interface introduced in this thesis was designed and implemented by myself on top of the ViDex platform [23] with permission. The research question, experiment designs, and data analysis came from my own efforts with the help of Dr. Ido Roll, Matthew Fong, Dr. Sidney Fels, and Dr. Yan Liu. The Certificate Number of the Ethics Certificate was H13-01589.

Table of Contents

Abstract . . . . . **iii**

Lay Summary . . . . . **iv**

Preface . . . . . **v**

Table of Contents . . . . . **vi**

List of Tables . . . . . **ix**

List of Figures . . . . . **xi**

Glossary . . . . . **xiii**

Acknowledgments . . . . . **xiv**

1 Introduction **1**

1.1 Research Question 4

1.2 Contributions 5

2 Background and Related Work **7**

2.1 Locating Information in Video 7

2.2 Annotations 11

2.2.1 Text Annotations 11

2.2.2 Video Annotations 13

2.3 General Discussion 19

3 Video Annotations in Information-Seeking 21
3.1 Experiment Design and Procedure 22
3.1.1 Participants 22
3.1.2 Design 22
3.1.3 Task 24
3.1.4 Materials 25
3.1.5 Procedure 27
3.2 Interface Design 27
3.2.1 Overview 27
3.2.2 Video Player Page 28
3.2.3 Annotation Manager 32
3.3 Measures 33
3.3.1 Task Performance 33
3.3.2 Surveys 33
3.3.3 Interviews 34

4 Results and Discussion 35
4.1 Task Performance 35
4.1.1 Training Session 35
4.1.2 Formal Session 1 37
4.1.3 Formal Session 2 39
4.2 Annotations 40
4.3 Surveys 41
4.3.1 Pre-Experiment Survey 41
4.3.2 Session Survey 43
4.3.3 Post-Experiment Survey 44
4.4 Interviews 45
4.4.1 Annotations 46
4.4.2 The Annotation Process 47
4.4.3 Annotating Video vs. Annotating Text 50
4.4.4 The Use of The AM 51
4.5 Summary 52

5 General Discussion 54
5.1 Interpreting the Results 55
5.2 Limitations 58

6 Conclusion and Future Work 61

Bibliography 63

A Surveys 70
A.1 Pre-Experiment Survey 71
A.2 Session Survey 75
A.3 Post-Experiment Survey 77

B Searching Tasks 79
B.1 Quantum Computation 79
B.2 Fusion Power 83
B.3 The Brain 87
B.4 No AM Condition 90

C User Study Data 91

D R Code 92

E Power Analysis 93

List of Tables

Table 1.1 Feature summary of popular online learning platforms. 2

Table 3.1 Demographics summary for subjects. 23

Table 3.2 Video watched and AM conditions for each subject in the formal session 1. Video 1 was about the brain, and video 2 was about fusion power. For AM, "O" means the same tasks as in Appendix B.2 for video 2, and Appendix B.3 for video 1; "S" means the reversed AM conditions compared to "O", i.e., items not assigned with the AM in Appendix B.2 or B.3 were assigned with the AM. 24

Table 3.3 Video watched by each subject in the formal session 2. Video 1 was about the brain, and video 2 was about fusion power. The searching items were the same as in the formal session 1, but without assigned AM conditions, as exemplified in Appendix B.4. 24

Table 4.1 Numbers of items searched with the AM, annotated items, and annotations by subjects in session 2. 40

Table 4.2 Average numbers of annotations made by 16 subjects. 41

Table 4.3 Subjects' responses to pre-experiment survey questions. 42

Table 4.4 Subjects' responses to session survey questions (Likert scale, 1 to 5) of session 1 and 2. 44

Table 4.5 Subjects' responses to post-experiment survey questions (Likert scale, 1 to 5). 45

Table 4.6 Summary of subjects' interview responses regarding annotations and the annotation process. 46
Table 4.7 Numbers of subjects' interview responses regarding annotating video vs. annotating text and the use of the AM. 50

Table E.1 Means of subjects' task performances in session 1 in the pilot study (measured in seconds). 94
Table E.2 Parameters for power analysis on G*Power. 94

List of Figures

Figure 1.1 The video player page of Coursera. 3
Figure 1.2 The course page of Coursera. 4

Figure 2.1 TEMPO [50], by Nathan Prestopnik et al. 9
Figure 2.2 Hitchcock [26], by Andreas Girgensohn et al. 9
Figure 2.3 Micro-level and macro-level scaffolding [12], by Salome Cojean et al. (1) control, (2) microscaffolding, (3) macroscaffolding, and (4) two-level scaffolding conditions. 10
Figure 2.4 Artificial landmarks [59], by Md. Sami Uddin et al. Media player (a, b, c), PDF viewer (d, e, f). A, d: standard - with no landmarks; b, e: icon - augmented with abstract icons; c, f: thumbnail - augmented with extracted content as thumbnails. . 11
Figure 2.5 Types and applications of text annotations [49], by IIia Ovsiannikov et al. 12
Figure 2.6 Perceived importance of annotation features [49], by IIia Ovsiannikov et al. 12
Figure 2.7 Video annotations anchors and content [5], by Olivier Aubert et al. 13
Figure 2.8 ToolScape [36], by Juho Kim. 14
Figure 2.9 MediaNotes developed at Brigham Young University. 15
Figure 2.10 CLAS [44], by Negin Mirriahi et al. 16
Figure 2.11 Anvil [37], by Michael Kipp. 16
Figure 2.12 VITAL [46], by Frank Moretti et al. 17
Figure 2.13 Video Traces [54], by Amit Saxena et al. 18

Figure 2.14 DynamicSlides [34], by Hyeungshik et al. 19

Figure 3.1 Experiment design. 23
Figure 3.2 Video player page. 28
Figure 3.3 Course page with annotation manager. 28
Figure 3.4 The window to add annotations. 29
Figure 3.5 Select a tag or create a new tag. 30
Figure 3.6 Hover over an annotation icon in the transcript section to see the details of annotations. 30
Figure 3.7 The filmstrip section. 31
Figure 3.8 The timeline of the video player section. 31
Figure 3.9 Play highlighted parts only. 32
Figure 3.10 Annotation manager. 32

Figure 4.1 Subjects' task performances in the training session. 36
Figure 4.2 Subjects' task performances in formal session 1. 38
Figure 4.3 Comparing numbers of items searched with the AM, annotated items, and annotations made by subjects. 40

Glossary

AM annotation manager

ANOVA analysis of variance

LMEM linear mixed effects model

M mean

SD standard deviation

Acknowledgments

I'd like to thank everyone who has helped me academically during the process of finishing this research project. Thanks to my supervisor, Dr. Sidney Fels, for overseeing this work, and Dr. Ido Roll, Dr. Yan Liu, Kyoungwon, Matthew, and Samuel for sharing with me their expertise and invaluable advice. Thanks to colleagues in the Human Communication Technologies Lab and the ViDex team for all of the stimulating discussions and generous help.

Chapter 1

Introduction

Today, millions of learners enjoy online learning on a number of platforms [3], and these platforms continue to evolve to improve learner experiences and learning outcomes. Despite all the advantages of online video-based learning, it still faces many design challenges, including the highly important question of how to engage learners [16]. Learning engagement, defined as "the student's psychological investment in and effort directed toward learning, understanding, or mastering the knowledge, skills, or crafts that academic work is intended to promote" [47], has been found to be positively correlated with the use of learning technology and learning outcomes [22]. Learner-content interactions are also found to be critical in ensuring satisfying educational experiences [4]. However, as summarized in Table 1.1, currently most major video-based online learning platforms support few learner-video interactions and create a highly passive learning environment.

For example, Coursera provides both the video and its transcript to learners on its [1]video player page, as shown in Figure 1.1. Moreover, the interface provides annotation features and a list of annotated transcript segments along with associated notes as well, which make Coursera the most interactive learning environment compared to other platforms listed in Table 1.1. However, the interface supports limited annotation features, only including highlighting and adding notes to transcript segments, and there is only one color available for these highlights; while

[1]https://www.coursera.org/learn/human-computer-interaction/lecture/s4rFQ/the-interaction-design-specialization

	Video	Transcript	Transcript navigation	Bookmark	Bookmark list	Note	Note list
Coursera	√	√	√	√	√	√	
Edx	√	√	√				
iversity	√						
Khan Academy	√	√	√				
Lynda	√	√	√	√	√		
Udemy	√			√	√		
YouTube	√						

Table 1.1: Feature summary of popular online learning platforms.

there is a list of annotated transcript segments and annotations (notes) on the video player page, no indication of annotations is shown outside of this page, as shown on its [2]course page in Figure 1.2. This point is especially important, as rewatching a video is common, where the first step is to select a video. Showing learners their annotations of a video outside of the video player page has the potential to help select a video to rewatch, so the learner does not need to click and watch every video to find the correct one for review.

Though videos are used pervasively for learning both in schools and at homes, very few learner-content interactions are provided to learners. A typical video mainly includes visual, audio, and textual elements, among which the audio and textual are highly overlapped. Currently, for example in most video players such as QuickTime Player or on YouTube, the learner is not well supported to utilize these three representations of information: to examine different parts of the visual, the only thing the learner can do is click on the timeline or press the "fast forward" or "play back" button; the learner can access the textual only if the transcript or captioning of the video is provided, and the only interactions available are clicking on the timeline to see captions and scrolling up and down to read the transcript. We see that when compared to reading books, watching videos for the purpose of learning is not well supported. While reading a book, the learner has a table of contents to jump to chapters or sections of interest, an index to find specific detail, a highlighter to mark sentences, a pen or pencil to write notes, and a bookmark to keep the reader's place so that they can return and continue reading easily.

Thus, richer learner-content interactions and a better use of video annotations for rewatching videos need to be investigated to improve online learning experiences.

[2]https://www.coursera.org/learn/human-computer-interaction/home/week/1

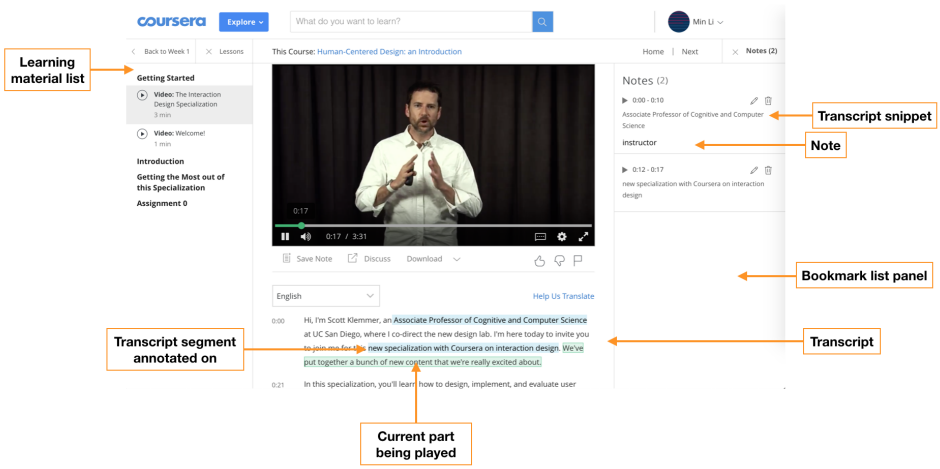


Figure 1.1: The video player page of Coursera.

Among the interactions with learning materials, annotations are particularly useful as they can be used by learners to think, to remember, and to clarify [49]. Annotating text has been performed by many learners for a long time and has also been explored extensively by previous research studies [49, 52, 57]. Though there are a number of studies exploring video interface designs with annotation features, few of them have explored the effectiveness of these features, and none of them have investigated the use of video annotations in locating in-video information such as video segments.

Rewatching is common in video-based learning, just as rereading textbooks is in classroom-based learning [8, 14]. To revisit a previously seen video segment, finding the location of the segment in the video is the first step. Unlike in a textbook, which has a table of contents and an index to support the learner in the information-seeking process, clicking on the timeline of a video is the only way to find a previously seen video segment. Though speeding up and fast forwarding are provided by most video players, the searching process still requires tedious trial and error especially for long videos.

The presented work explores learner-created annotations on the integration of video and text, and hopes to find whether video annotations can help learners locate previously seen video segments. In the proposed design, both the video and

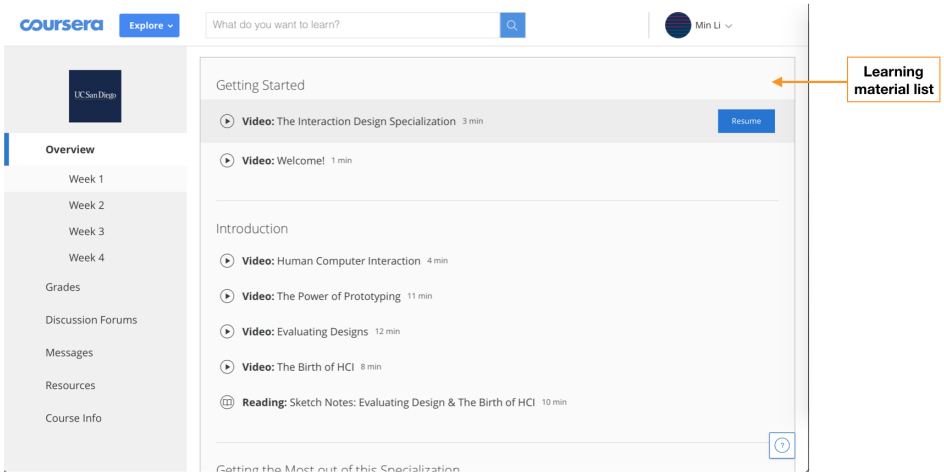


Figure 1.2: The course page of Coursera.

its transcript are provided and synchronized. This design is expected to improve the video-based learning experience and further support learners in the process of finding segments to review.

1.1 Research Question

This work explores the use of learner-created video annotations in locating previously seen information in a video. A research question that directs the entire research process is proposed:

> How effective are video annotations, which are created by learners while learning with both visual representation and textual representation of a video, in helping learners locate in-video information for revisitation?

Two hypotheses are also proposed:

1. Annotated information can be located faster than non-annotated information.
2. The more annotations the learner has made, the more likely they would rely on annotations to locate information.

1.2 Contributions

The contributions of the presented work consist of the design of a novel annotation interface which provides quick accesses to annotated video segments, and the results of a controlled lab experiment. In summary, evidence has been collected from subjects on:

1. the effectiveness of utilizing video annotations to locate previously seen video segments;

2. the effectiveness of the novel design of annotation manager in helping learners find previously seen video segments;

3. the effectiveness and limitations of the integration of text with video for video-based learning.

There are two main components in the research question: locating in-video information for revisitation, and learner-created video annotations based on the integration of a video's visual representation with textual representation. To answer this, a novel interface for video annotations was proposed and a controlled lab experiment investigating in-video information-seeking was carried out. The proposed interface consists of two components: an annotation mechanism based on an integration of text with video, and an annotation manager which enables the learner to see all annotations he/she has made on a video and provides quick accesses to annotated video segments. In the controlled lab experiment, tasks of locating in-video information were completed after the subject finished learning with and annotating a video using the proposed interface.

Various data were collected to answer the research question and understand related issues. Time spent on locating in-video information was analyzed to find whether annotated information was located faster; annotations made by subjects were also collected and analyzed to understand subjects' preferences regarding different types of annotations; qualitative data were collected and analyzed to better understand how subjects perceived video annotations, the annotation process, the proposed annotation manager, and the difference between annotating video vs. text. All of these data were found later on to complement each other and provide a richer answer to the research question.

Previous work on locating in-video information and video annotations are reviewed in Chapter 2. Chapter 3 describes the experiment design and the interface designed for the experiment. Experiment results are presented and discussed based on category in Chapter 4. General discussion which integrates results of all categories and the limitations of the presented work are given in Chapter 5. Finally, Chapter 6 gives conclusions of the presented work and proposes directions for future work.

Chapter 2

Background and Related Work

As the presented work explores how in-video information can be better located, this chapter reviews previous work on in-video information-seeking in section 2.1, to further explore the problem and identify gaps that the presented work can fill. The presented work utilizes video annotations, inspired by text annotations, to help learners locate in-video information for revisitation, and proposes an interface for creating and managing video annotations. Thus, related works on text annotations and interface designs for video annotations are reviewed in Section 2.2. Finally, a discussion and conclusion is made in Section 2.3 to emphasize the gaps in literature that the proposed work can fill.

2.1 Locating Information in Video

Unlike traditional classroom-based learning, learning with videos provides easier ways to learn with the same material over and over again. Rewatching has been found to be common in video watching experiences with 92% of viewers having rewatched some type of video every month, and rewatching tutorial videos several times was perceived as necessary for learners to finish related tasks [8]. Online learners were found to rewatch parts of video lectures frequently [14], and rewatching segments instead of the whole video was also found to be more helpful in learning [42].

To rewatch a video segment, the learner needs to first find the respective seg-

ment, which has been found to be not easy. Undergraduate students were found to search for in-video information using low level strategies and focused more on local information both temporally and spatially, lacking attention to global patterns [39]. In the experiment, Richard Lowe asked subjects to learn with an interactive animation and apply what they learnt to predict and draw the pattern of meteorological markings. Video-like controls were provided in the animation. Results showed that subjects revisited parts of the animation frequently, but used low level strategies that addressed limited temporal and spatial scopes. The searching process was more a simple off-loading of information rather than conceptual grouping of various parts. It is worth noting that all subjects of this experiment were novices in meteorology, so learners with more expertise may perform differently in the tasks. The animation used in the experiment did not have audio which may have made the searching process more complicated, but indeed more similar to the real life case of searching for information from videos.

The results of Richard Lowe's experiment are valuable in understanding the process of learners assimilating and applying new knowledge from videos, and they indicate the need to better equip learners to revisit in-video information. Recently, a more advanced interface was designed to better visualize the temporal relationships among events happening in the same locations. As shown in Figure 2.1, the Temporally Enabled Map for Presentation (TEMPO) interface indicated critical events on the timeline, after the designers found that viewers weighed moments with various importance [50]. The indications may serve as an implicit summary or table of contents of the animation, which help the learner locate points of key information quickly, and thus does not require tedious examination of points to find the one of interest. However, how those indications can actually help in the information-seeking process is yet to be experimented.

Locating video segments in a large video database can be taxing and time-consuming using low level strategies such as browsing through, as the user will have to watch videos in their entirety. To address this problem, many video summary techniques have been invented to extract key video frames or create previews automatically. For example, the Hitchcock system shown in Figure 2.2 is able to show the user a preview with key frames, and also enable the user to search video segments based on text information of the video, such as its transcript [26],

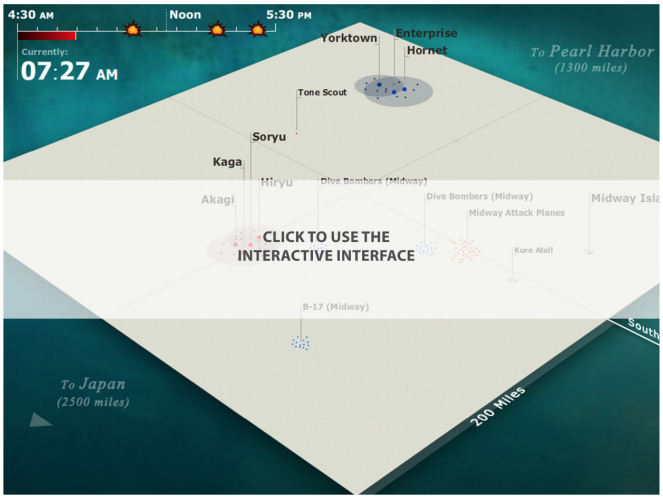


Figure 2.1: TEMPO [50], by Nathan Prestopnik et al.

Figure 2.2: Hitchcock [26], by Andreas Girgensohn et al.

though the effectiveness has not been examined. The Visual Transcript combines the visuals and the transcript segments of a lecture video to create a document style summary, and the text component of the summary was found to be helpful to learners in finding information [55].

Another stream of research explores interface designs to scaffold users in the process of locating video segments on their own. Recently, Salome et al. found positive effects of both micro-level and macro-level scaffolding on information-seeking in video-based environments [12, 13]. In their experiments, macro-scaffolding was provided by a table of contents of the video, and micro-scaffolding was realized by segmentation markers on the timeline, as shown in Figure 2.3. Those scaf-

Figure 2.3: Micro-level and macro-level scaffolding [12], by Salome Cojean et al. (1) control, (2) microscaffolding, (3) macroscaffolding, and (4) two-level scaffolding conditions.

foldings increased the saliency of key information and also created external mental models of the video content. Thus learners could find information faster with this external mental model provided on the interface.

Except for the focus on video content, spatial memory has also been considered as important in locating information in video for revisitation. For example, artificial landmarks, as shown in Figure 2.4, were found to be helpful in building up spatial knowledge of a video so that learners could navigate back to locations visited before more easily, and arbitrary icons were found less effective than thumbnails which were customized for each video [45, 58, 59].

All of these recent efforts on locating in-video information summarized above investigated the use of provided information instead of users' own inputs, such as annotations. Though these interactive features have been found effective in improving learning outcomes generally [17, 66], which will be introduced shortly, their effectiveness in the information-seeking process has been left behind. To fill this gap, the proposed work explores how effective learners' self-entered annotations are in helping locate in-video information.

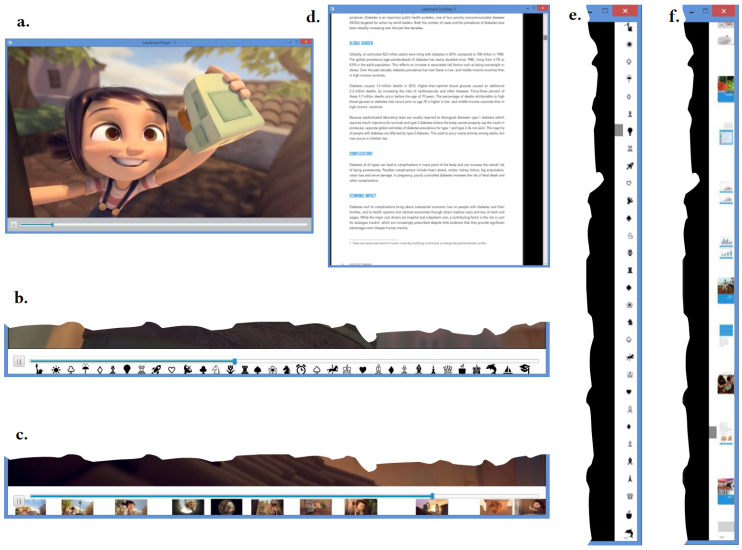


Figure 2.4: Artificial landmarks [59], by Md. Sami Uddin et al. Media player (a, b, c), PDF viewer (d, e, f). A, d: standard - with no landmarks; b, e: icon - augmented with abstract icons; c, f: thumbnail - augmented with extracted content as thumbnails.

2.2 Annotations

This section takes a closer look at annotations. Text annotations are discussed first, followed by an examination of interface designs for video annotations.

2.2.1 Text Annotations

As the proposed design for video annotations was inspired by text annotations, it is worthwhile to gain an understanding of text annotations. Discussions of the results of the proposed experiment will refer back to studies examined in this section.

Active reading is a process of assimilating and reusing the reading material as part of the reader's knowledge network [62], and has been considered as a combination of reading, thinking, and adding annotations [1]. With the development of technology, annotating documents is more supported and popular than before. Today's document management software, such as Endnote and RefWorks, provide annotation tools for PDF documents, but they rely heavily on metadata and require

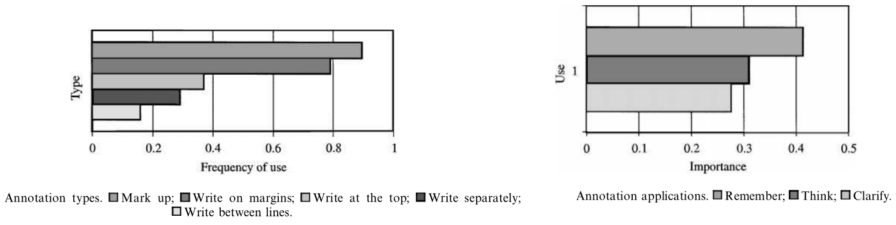


Figure 2.5: Types and applications of text annotations [49], by IIia Ovsiannikov et al.

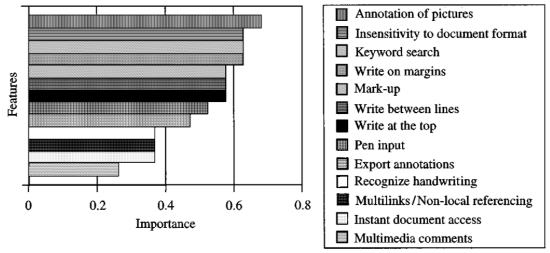


Figure 2.6: Perceived importance of annotation features [49], by IIia Ovsiannikov et al.

a high level of proficiency. Interestingly, there were more discussions regarding annotations for electronic files in the 1990s. For example, DynaText supports three types of annotations: bookmarks, notes, and hyperlinks [56]. The DynaText annotation manager supports sorting, viewing, and deleting. Directing users from annotations to corresponding pages was implemented by another annotation system, Re:mark. Comment sharing and collaborative commenting have also been developed [15, 38].

Though advocates of e-reading claim that e-reading provides more advanced features to support learning and improve learning outcomes [52], many students still prefer to read on paper, and navigation and annotation functions of e-reading have been found to be inferior to paper [57].

IIia Ovsiannikov et al. asked readers 3 sets of questions targeting primarily the research and academic environment, based on which they identified main annotation types, major annotation applications, and perceived importance of annotation

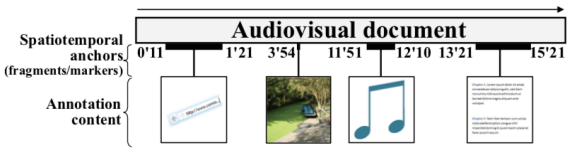


Figure 2.7: Video annotations anchors and content [5], by Olivier Aubert et al.

software features [49]. Those readers were undergraduate students, graduate students, professors, and professionals. Regarding locations of annotations, readers usually mark up text segments, write on margins, write at the top, write in a separate document, or write between lines. The frequency of use of each is shown in Figure 2.5. These annotations were used mainly for three purposes: to remember, to think, and to clarify. Regarding electronic annotation features, the perceived importance of features varied, as shown in Figure 2.6.

2.2.2 Video Annotations

Just as text annotations, two main components of video annotations are usually of interest: the content and the anchor.

For a text annotation, the anchor is in the spatial dimension of the text; however, for a video annotation, both temporal and spatial dimensions are included, as illustrated in Figure 2.7. For example, the video timeline is usually used to address the temporal dimension of video annotations. The spatial dimension is usually simplified as the whole video frame. For example, the video frame at the time point when an annotation is created is associated with the annotation. Few designs have utilized specific spatial regions of video frames to make annotations [55, 61]. Video Annotation Learning System allows users to add graphical annotations such as arrows onto videos, and lists the types of annotations with timestamps [11].

The integration of both temporal and spatial dimensions of video annotations can make it difficult to display information. As shown in Figure 2.8, the limited region around the video timeline provides a very crowded space to display thumbnails with simple annotations as marks [36]. In the proposed interface, only the temporal dimension is addressed by displaying marks of annotations in corre-

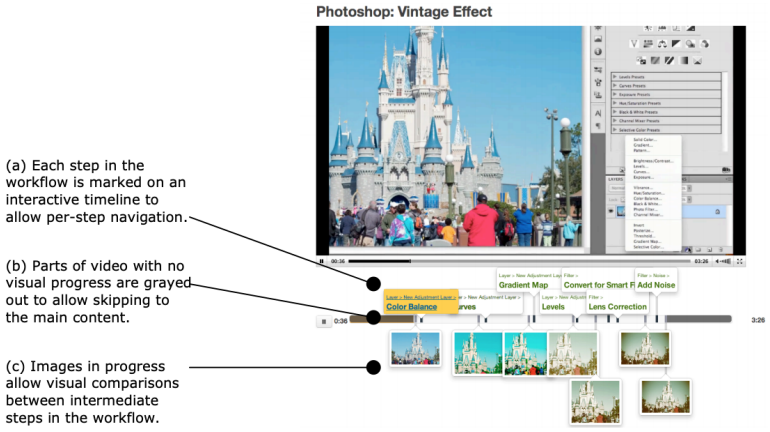


Figure 2.8: ToolScape [36], by Juho Kim.

sponding locations of the timeline, the transcript of the video, and the filmstrip.

When users are allowed to type in their own notes, there is usually a note list provided [32, 51, 63]. As shown in Figure 2.9, MediaNotes[1], which was developed at Brigham Young University, has a highly comprehensive set of annotation features, including naming (section A), segmenting (section B), commenting (section C), and tagging (section D). Annotations are grouped based on timeline positions of annotated segments in a panel alongside the video player. Annotations can be used for complex data mining such as filtering, analysis across time, space, tag set, or person. In the proposed design, annotations are grouped into lists as well; in addition, the proposed design enables linking to specific video segments from annotations and allows learners to review their annotations outside of the video where there may be more than one video, increasing convenience for the purpose of reviewing as the learner does not need to play each video again.

Aside from a list, video annotations can be well visualized in the video timeline. In the Collaborative Lecture Annotation System (CLAS) [44], users can comment on a specific point of the video or the whole video, and annotations of the video are mapped to the timeline in a separate panel below the video player. Annotations of the users who watched the same lecture video are displayed and analyzed collectively in this system, and the instructor of the course can see an analysis of

[1] https://www.youtube.com/watch?v=DmrlV99FL3E

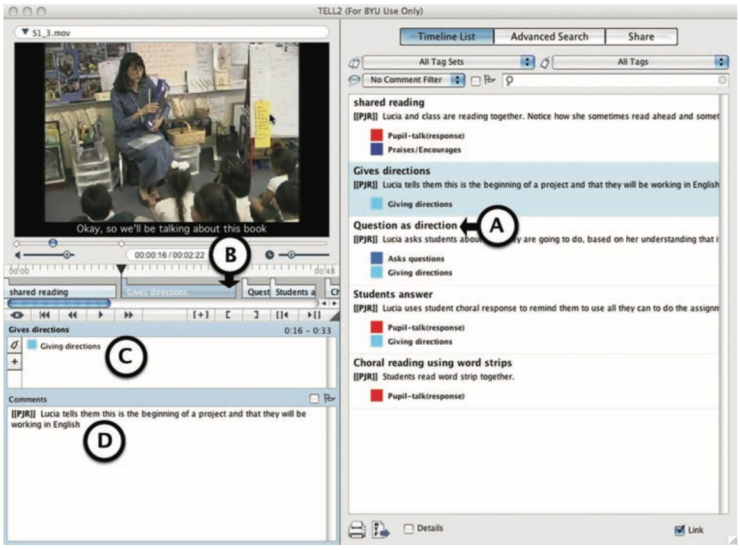


Figure 2.9: MediaNotes developed at Brigham Young University.

student video watching activities, as shown in Figure 2.10.

There have been many video annotation systems designed for data analysis, and these systems display annotations in a similar way [21, 35, 37, 68]. For example, Anvil displays annotations in parallel tracks based on themes used in the data coding process by the user [37], as shown in Figure 2.11. Observer XT used a similar visualizing technique, but with multimodal data such as behavioural data, sound data, and physiological data, all in the form of text [68]. These systems usually emphasize annotations themselves, for example codes, and the integration of annotations with the original material (such as the transcript) is not well supported.

Video annotations along with corresponding video clips can also be used to write essays, and they appear as hyperlinks in this case [7, 46]. In the Video Interactions for Teaching and Learning (VITAL) system, users can add notes to video segments as they watch videos, and embed annotated video segments as video-based evidence in essays. As shown in Figure 2.12, annotated video segments are listed in the left panel, from which users can embed them into the essay on the right. Readers of the essay can view the video-based evidence by clicking on the hyperlink while reading.

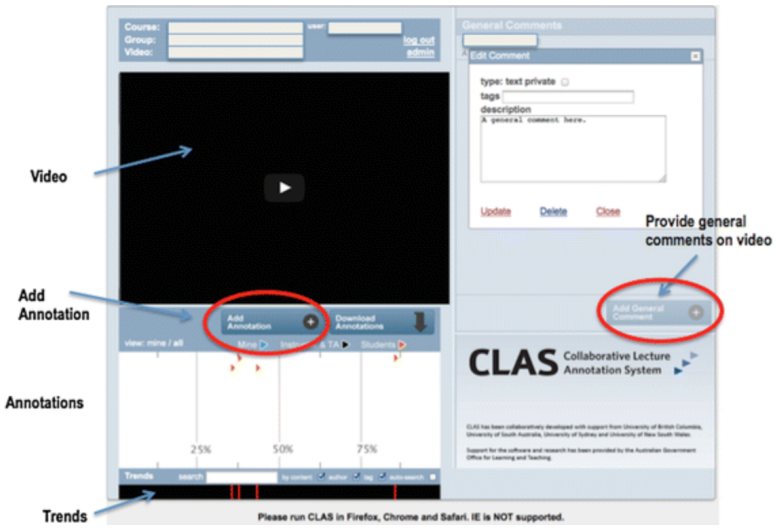


Figure 2.10: CLAS [44], by Negin Mirriahi et al.

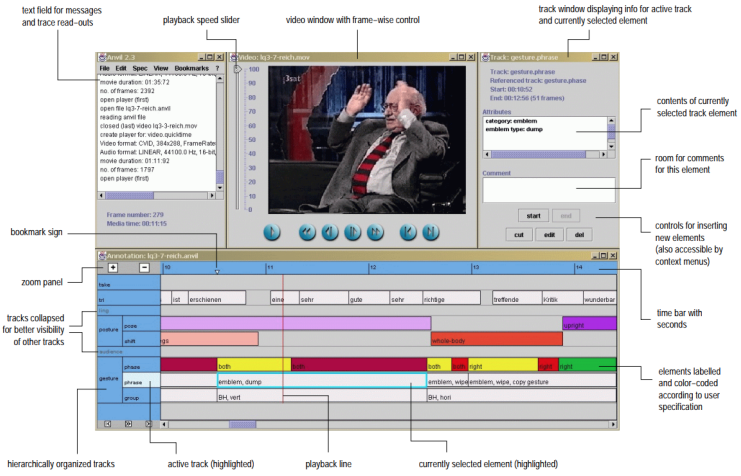


Figure 2.11: Anvil [37], by Michael Kipp.

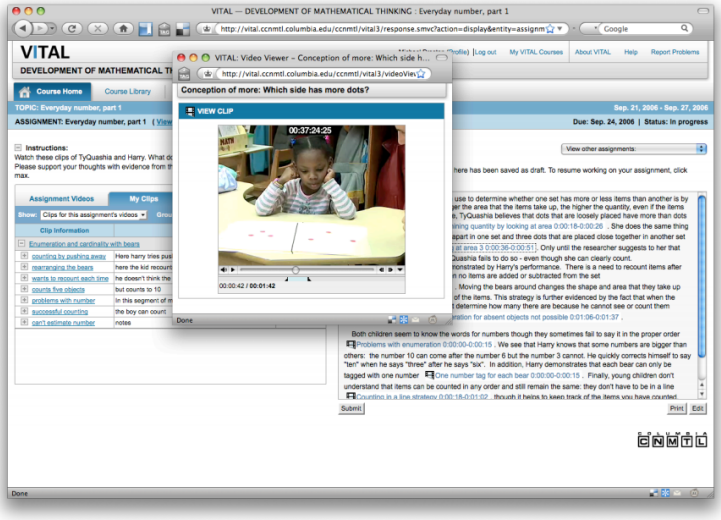


Figure 2.12: VITAL [46], by Frank Moretti et al.

Though text is the most common form of video annotations, audio and visual annotations have been used in some systems as well. VideoPaper offers users the option to add images as video annotations [7]. In Video Traces [54], the definition of annotations is broader than other systems discussed above: users can annotate voice, pointing, and drawing to images and audio-video files, as shown in Figure 2.13. Annotations in this system are better used to stimulate collaborative learning, student-student interaction, and learner-teacher interaction as users can annotate on the same video collaboratively and can also respond to annotations by creating a "threaded discussion". The threaded discussion is similar to the annotation manager in the presented work, except that the threaded discussion aggregates annotations of different learners and displays different types of annotations in one place.

Recently, researchers have tried to create annotations automatically for learners based on eye-gaze tracking. In GazeNoter which was developed in 2016, notes are taken automatically based on the viewer's eye-gaze patterns. The video content relevant to the current notetaking is highlighted, so the user is fully aware of the notetaking process. When the user types in notes manually, video playback is slowed down; the video is paused if the video is about to change to another slide

Figure 2.13: Video Traces [54], by Amit Saxena et al.

when the user is taking notes manually. This design was reported by all participants as helpful for them to focus and think while writing [48].

While the annotation process of most video annotation systems relies on visual information, there have been some utilizations of video transcripts at different levels, among which systems for qualitative data analysis account for a big proportion because of the data coding process [21, 37, 41, 60]. The audio channel of educational videos contains as much as, if not more, information as teachers talking in classrooms. Recently, DynamicSlide [34] has displayed the transcript of a video on the right side of the video player as individual segments which are synchronized with the video, and in-video objects were linked to corresponding transcript segments. As shown in Figure 2.14, the system enables users to make annotations on in-video objects, and notes are listed below the video player where learners can play the corresponding video segment.

In summary, a number of video annotation systems have been designed for various applications, but none of them provide access to annotations of multiple videos in the same location. In the proposed design of the annotation manager, the learner is able to see annotations of multiple videos on the same page (the course

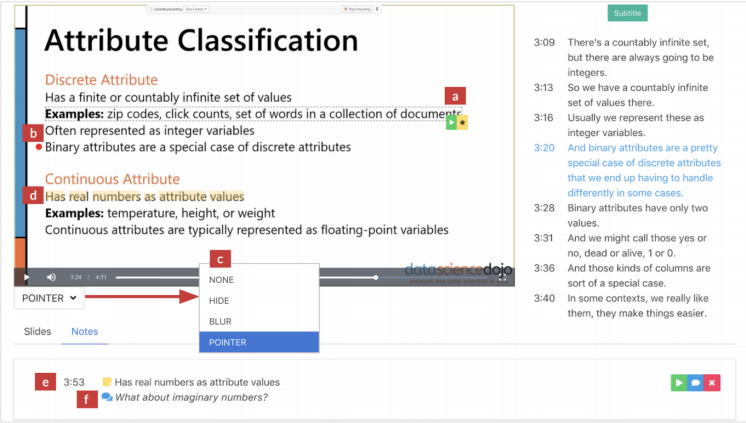


Figure 2.14: DynamicSlides [34], by Hyeungshik et al.

page), and these annotations act as hyperlinks that direct the learner to corresponding video segments, offering a more convenient way to locate and rewatch video segments.

2.3 General Discussion

Electronic devices provide more options for interactivity in processing both text and video information, and interactivity has been found helpful for both types of information processing. Though some video-based learning systems provide transcript along with video, the full potential of video transcripts in learning is yet to be unlocked.

Video annotations, as an important type of interactivity in video, are distinct from text annotations due to the unique nature of video; though their effectiveness in improving learning outcomes has been confirmed, there has been little evidence for how they can support learners in the process of locating in-video information, which is critical for revisiting video segments. Previous studies have suggested that learners did not make many video annotations [18–20, 30], but it is worth noting that subjects in those studies tended to annotate on video content that they wanted to revisit in the future. Further, platforms used in these studies did not have features supporting learners in this revisitation through video annotations. Thus, it is possible that learners did not annotate in larger quantities due to the absence of

a mechanism supporting the use of video annotations in revisiting video content of interest. Moreover, though video annotations were found to be used as bookmarks for future rewatching, how much they can help in locating video content for this rewatching remains untapped.

The two major parts of the presented work are video annotations and the use of them for in-video information searching. The presented work creates a closed loop video annotation process from creating annotations to the use of them after creation, and thus fills present gaps in both video annotations and locating in-video information.

Chapter 3

Video Annotations in Information-Seeking

This chapter presents the methods used to answer the research question posed by this study. In the presented work, an interface was designed and implemented as the apparatus, and an experiment was designed and carried out to collect data based on several measures.

In the proposed interface, video annotations are displayed in two places: the AM which collects all annotations made on the video, and the video player page where video annotations are displayed along with the original materials, i.e., the video, the transcript, and the filmstrip. Thus, there were two ways of using video annotations to locate information: through the AM, and on the video player page. To investigate whether the AM helped in the information-locating process, subjects were asked to finish searching tasks under both the AM condition and the non-AM condition; locating annotated vs. non-annotated information was also compared to answer the research question. Searching time was analyzed to determine whether the AM helped in reducing searching time, and whether annotated information was located faster. To provide a richer answer to the research question, and to better understand the components of the research question, annotations made by subjects and responses to survey and interview questions were also collected.

3.1 Experiment Design and Procedure

3.1.1 Participants

The interface targeted at undergraduate students, and the presented work includes both quantitative study and qualitative study. The quality of the quantitative study was ensured by assigning subjects to experiment conditions randomly, which will be introduced shortly in this chapter, and a power analysis based on data from a pilot study. To ensure the quality of the qualitative study, UBC undergraduate students with diverse academic backgrounds were recruited as subjects of the presented experiment through an online posting; the number of subjects was also determined so that data saturation could be reached.

16 subjects were recruited and this number was in line with previous studies on similar topics[2], and was also much higher than the number 4, which was suggested by the power analysis using data from a pilot study, a simplified version of the formal experiment. Details of the power analysis can be found in Appendix E.

Subjects were between 18 to 27 years old, and the mean age was 20.63. They were from diverse academic backgrounds, including arts, science, engineering, and commerce; the numbers of participants in each academic year were balanced, with 5 in the first year, 3 in the second year, 3 in the third year, and 5 in the final year. Subjects' basic information can be found in Table 3.1. Subjects' prior experiences with the ViDex platform, video watching, video rewatching, note-taking for videos, and annotating texts are summarized in Section 4.3.1 based on survey responses.

3.1.2 Design

The experiment design and the procedure for each subject are shown in Figure 3.1. The mini session was used to get subjects familiar with the interface and its features. Following the mini session, there were three sessions in the experiment: the training session, the formal session 1, and the formal session 2. Subjects in early pilot studies reported that they found their strategies after finishing one session, so the training session was used for subjects to develop their strategy of learning, annotating, and information-seeking.

Subject ♯	Gender	Age	Major	Year
1	F	19	Math	2
2	F	19	Geological engineering	3
3	M	23	Psychology	4
4	F	19	Food, nutrition, and health	2
5	M	21	Physiology	4
6	M	27	Life science	4
7	F	19	Arts	1
8	F	24	Law	1
9	M	19	Mining engineering	1
10	F	21	Physics	3
11	M	19	Commerce	1
12	M	18	Science	1
13	F	20	Food, nutrition, and health	3
14	M	21	Math	4
15	F	20	Anthropology	2
16	F	21	Commerce	4

Table 3.1: Demographics summary for subjects.

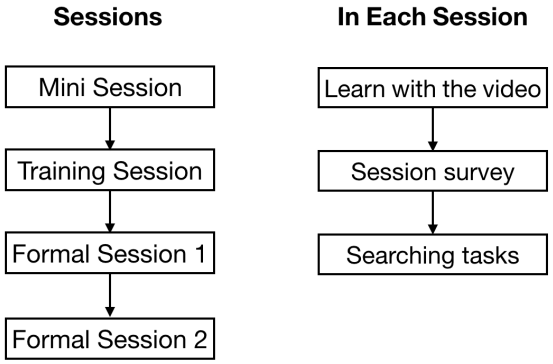


Figure 3.1: Experiment design.

The formal session 1 investigated how the AM affected subjects' task performances and whether annotated information was located faster than non-annotated information. Thus, subjects were asked to finish half the searching tasks with the annotation manager (AM), and the other half without the AM. The formal session 2 was to find whether subjects would use the AM and annotation features spontaneously, so subjects were not required to use the AM for any tasks; whether to use the AM or not was their own choice.

In the training session and the two formal sessions, subjects watched a video at their own pace, and they were free to use any features on the video player page. They then filled a survey which was the same for all sessions; finally, they completed a set of 12 searching tasks, which will be described shortly. The number of tasks in each session was determined by the time constraint, as the experiment

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
Video	1	1	2	2	1	1	2	1	1	2	1	2	2	2	1	2
AM	O	S	O	S	S	O	O	S	O	S	S	O	S	O	O	S

Table 3.2: Video watched and AM conditions for each subject in the formal session 1. Video 1 was about the brain, and video 2 was about fusion power. For AM, "O" means the same tasks as in Appendix B.2 for video 2, and Appendix B.3 for video 1; "S" means the reversed AM conditions compared to "O", i.e., items not assigned with the AM in Appendix B.2 or B.3 were assigned with the AM.

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
Video	2	2	1	1	2	2	1	2	2	1	2	1	1	1	2	1

Table 3.3: Video watched by each subject in the formal session 2. Video 1 was about the brain, and video 2 was about fusion power. The searching items were the same as in the formal session 1, but without assigned AM conditions, as exemplified in Appendix B.4.

aimed at finishing within 1.5 hours.

All subjects watched the same video and finished the same set of searching tasks in the training session, while the orders of the other two videos for the two formal sessions were counterbalanced so each video was watched by half of the subjects. To mitigate the effect of the videos on subjects' task performances, the order of videos in the two formal sessions was counterbalanced. Though the order of tasks for each video was fixed, the AM conditions associated with tasks were counterbalanced, as some tasks may be intrinsically easier to locate with the AM. Due to this experiment design, simple within-subjects or between-subjects analysis were not applicable to the data obtained in the presented work. Thus, the linear mixed effects model (LMEM) was used for analysis, which will be introduced in Section 4.1.

Counterbalancing for videos and AM conditions (described in the following section) in the formal session 1 is summarized in Table 3.2. Counterbalancing for videos in the formal session 2 is shown in Table 3.3.

3.1.3 Task

There were twelve searching tasks in each session. In the training session and the first formal session, half of the searching tasks asked the subject to finish the

task with the AM, while the other half without the AM; the order of searching tasks of a video remained unchanged, but the AM conditions assigned to tasks were counterbalanced in the formal session 1, as shown in Table 3.2. In the formal session 2, the subject was free to use the AM at his/her own discretion, i.e., there was no AM condition assigned to each task and the subject decided whether to use the AM or not on his/her own.

The goal of a searching task is to locate a piece of information in the original material, a video in the case of the presented experiment. For each searching task, the subject started from the course page and completed the task either with or without the AM. When using the AM, the subject entered the video player page by clicking on an annotation in the AM; the subject searched through the AM of the video to find the annotation that he/she thought would be the most helpful to find the to-be-located information on the video player page. When not using the AM, the subject entered the video player page by clicking on the cover frame of the video on the course page. After the subject entered the video player page, he/she was free to use any features on the page to find the point of the target information. If the subject found the correct point, the experimenter would ask the subject to go back to the course page, indicating the completion of the task. There were no time constraints for tasks, but subjects were asked to try their best to finish each task as quickly as possible.

3.1.4 Materials

Videos

Early small-scale pilot studies showed the following: subjects became bored or tired in the middle when watching a 13-minute long video; if the video was too easy, subjects learned passively, ignoring the video transcript and not making any annotations; if the video content was too difficult, subjects became lost, especially when they did not have enough prior knowledge; if the visuals or the transcript of a video contained little information, subjects tended to ignore them and focus on the other presentations of video content. In light of the above, each of the three videos used in the experiment were (a) 5 to 7 minutes long, (b) at medium to high difficulty

levels (rated as 3-4 on a 1-5 difficulty scale in pilot studies), (c) intense information covered both visually and verbally (the speaker in the video spoke continuously, and the video frame changed for the most part within every 10 seconds), and (d) an introduction to the topic that did not require much prior knowledge (each video has been viewed for more than 2 million times and received various comments on YouTube, which indicates that viewers with various backgrounds were able to understand the video in some way).

Videos used in the experiment were 3 segments of popular YouTube videos designed to educate the public. The video used in the training session was about [1]quantum computation, and the other two videos were about [2]the brain and [3]fusion power respectively.

To-be-located Information

A piece of to-be-located information was considered as a searching item. Searching items for all three videos are shown in Appendix B. Appendix B.1 is the set of searching items for the training session, which was the same for all subjects. Appendix B.2 and B.3 show what the searching items for the two videos looked like in the formal session 1 (the counterbalancing of AM conditions has been explained in Section 3.1.2). Appendix B.4 is an example of how subjects were reminded that there was no AM condition assigned, and that they were free to either use or not use the AM.

For the searching items of each video, half of them were summarizations of segments of the video content, thus not only memorizing but also understanding of the video content was required for the learner to locate them, which simulated the real-life case. Each topic was ensured to appear in the video only once. To simulate cases when various types of information remind the learner which content to revisit, using the other half of topics, three of them were original transcript segments and the remaining three were video frames. To prevent predicting the location of the to-be-located information, and to ensure the independence of each task, the order of these information pieces was randomized so that they did not follow the same

[1]https://www.youtube.com/watch?v=JhHMJCUmq28. Accessed on February 13, 2018.
[2]https://www.youtube.com/watch?v=kMKc8nfPATIt=473s. Accessed on February 13, 2018.
[3]https://www.youtube.com/watch?v=mZsaaturR6E. Accessed on February 13, 2018.

chronological order as in the original material.

Each piece of to-be-located information lasted for a short period of time in the video, thus a task was considered complete if the subject located the information at any point within the corresponding time range.

3.1.5 Procedure

After a short greeting, the subject was asked to sign a consent form, then to complete the pre-experiment survey. Next, a verbal introduction as well as demonstrations of how each feature worked were given by the experimenter, followed by a mini session to allow the subject to fully experience the interface. In the mini session, the subject watched the first two minutes of the training session video, and was asked to add six annotations (two for each type) and perform six searching tasks including all possible task types. Then the subject went through 3 sessions: the training session (also referred to as session 0), the formal session 1, and the formal session 2.

The procedure for each session was the same, as shown in Figure 3.1. In each session, the subject watched the video first, then filled out a survey, and finally performed twelve searching tasks. While watching the video, subjects were free to watch as many times as they wanted, and they could use any feature on the video player page. The survey used in each session was the same. After each session, the subject was offered the option to take a three-minute break.

After all three sessions, the subject was asked to complete the post-experiment survey, followed by a semi-structured interview.

The entire process lasted for about one and a half hours. The screen of the subjects finishing their searching tasks was recorded and used to analyze searching time.

3.2 Interface Design

3.2.1 Overview

The proposed interface consists of two main components, the video player page for learning and annotating, and the course page for course management and annota-

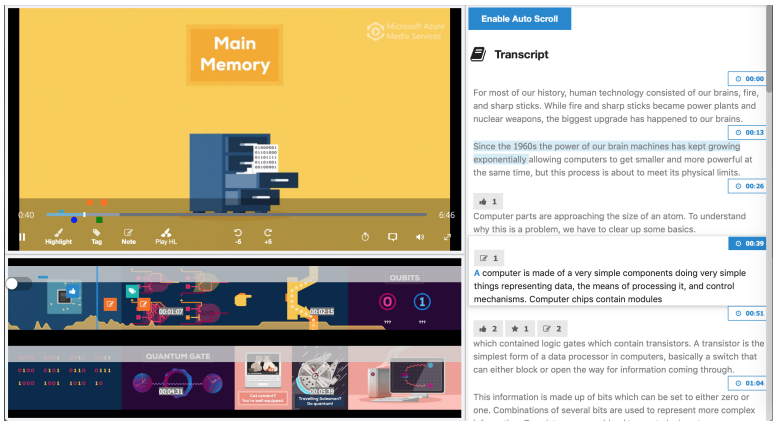


Figure 3.2: Video player page.

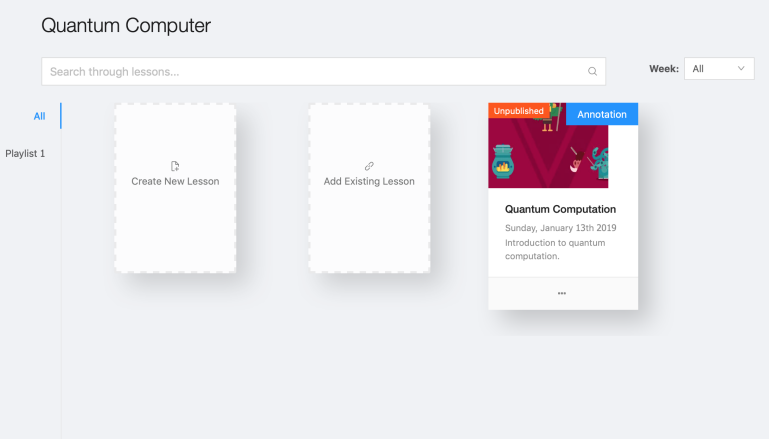


Figure 3.3: Course page with annotation manager.

tion management. The video player page aims at creating a seamless integration between video and text to fully utilize both the visual and audio information of the video. The course page collects all videos of a course and provides an annotation manager for each video.

3.2.2 Video Player Page

The video player page was adapted from the ViDex platform [23]. As shown in Figure 3.2, the video player page consists of three sections: a video player, a tran-

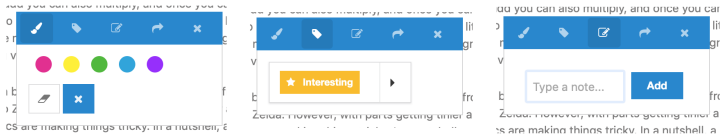


Figure 3.4: The window to add annotations.

script, and a filmstrip. Contents of these three sections are synchronized and are three presentations of the same learning material. Learners watch a video and make annotations on this page, and they can navigate the video by interacting with any of these three sections. Annotations and the progress of video watching are also displayed in all three sections, which are discussed briefly in following paragraphs. Annotations are synchronized with the video and are mapped to corresponding locations in all three sections.

There are three types of annotations that the learner can make: highlight, tag, and note. The learner can add annotations in the transcript section and the filmstrip section following the same procedure: select a video segment by clicking and dragging the mouse, and then select or create an annotation in the pop-up window. As shown in Figure 3.4, the window for adding annotations contains five icons: one for highlight, one for tag, one for note, one for sharing, and one to close the window. To add a highlight, the learner just needs to select a color; the learner can also delete a selected highlight by clicking the eraser button below the color selection region. To add a tag, the leaner needs to select a tag first, and then click on the selected tag to associate it with the selected segment. As shown in Figure 3.5, a list of existing tags is shown if the learner clicks on the arrow sign; a window for creating a new tag pops up if the learner clicks the “Add New Tag” button. To add a note, the learner simply fills in the blank by typing, and then clicks the “Add” button. If the learner hovers over any annotation icon in any of the three sections, he/she will see the details of the corresponding annotations, which will be described shortly in following paragraphs.

The transcript section is a place to display both the literal transcript of the video and annotations. The transcript is divided into segments with time stamps, and each segment is clickable so that the learner can jump to a specific video segment at will.

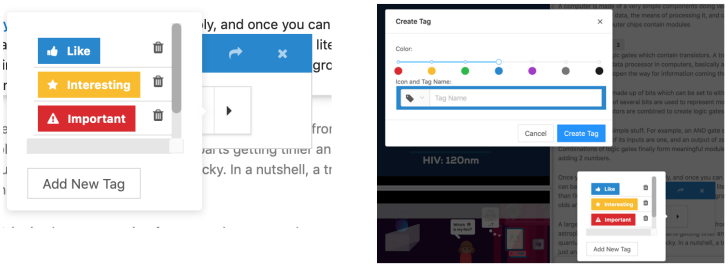


Figure 3.5: Select a tag or create a new tag.

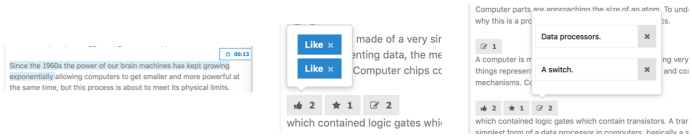


Figure 3.6: Hover over an annotation icon in the transcript section to see the details of annotations.

The current segment being played is indicated by the shadow around the transcript segment. Tags and notes made for each segment are shown by icons above the segment; the number following each icon indicates the number of the same type of annotations made for the segment. If the learner hovers over any icons above a transcript segment, he/she will see the details of the corresponding annotations, as shown in Figure 3.6. The learner can also delete tags and notes, or modify notes here.

The filmstrip, as shown in Figure 3.7, is a collection of video frames with equal time intervals between any two adjacent frames, along with indicators for annotations, a viewing history heat-map, and a vertical bar representing the current frame being played. The heat-map can be disabled by clicking the switch at the top-left corner. Highlights are indicated by color bars at the top; tags and notes are indicated by corresponding icons in the middle of the filmstrip. Icons for tags are placed at the same horizontal line above icons for notes, and adjacent tags or notes are grouped into one icon to avoid overlapping. If the learner hovers over an annotation icon, he/she will see the details of the annotations and modify or delete

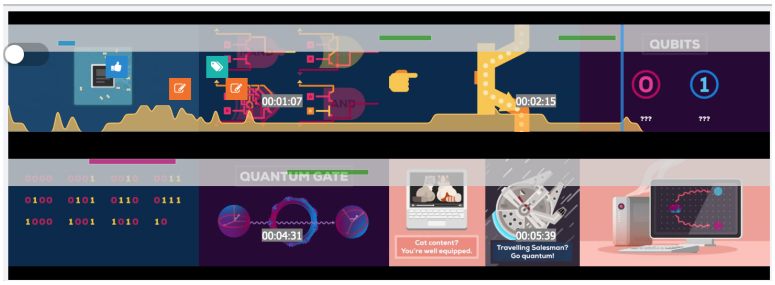


Figure 3.7: The filmstrip section.

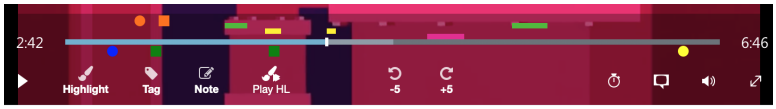


Figure 3.8: The timeline of the video player section.

annotations, just as in the transcript section. The learner can also click on any place of the filmstrip to play the video from a specific point at will.

In the video player section, video control options are provided, including a play/stop button, speed up and slow down buttons, a button to switch caption options, a button for volume adjusting, and a button to enter full-screen mode, as shown in Figure 3.8. Moreover, the learner is able to add annotations and play highlighted parts only. To add an annotation in this section, the learner simply clicks any of the three annotation-adding buttons below the timeline bar: "Highlight", "Tag", "Note". To play only the highlighted parts, the learner can hover over the "Play HL" button and then select a highlight color to play in a pop-up window, as shown in Figure 3.9. Annotations are displayed around the timeline bar at different horizontal lines depending on their types. Tags are displayed underneath the timeline bar as circles and squares representing a single tag and grouped tags, respectively; the colors of circles representing single tags are the same as the tags, while all squares representing grouped tags are green. Highlights are displayed as colored bars above the timeline bar, and the colors are the same as corresponding highlights. All notes are orange with single notes as circles and group notes as squares. The learner can also hover over indicators of tags or notes to see the details.

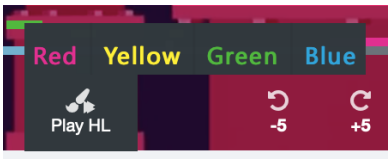


Figure 3.9: Play highlighted parts only.

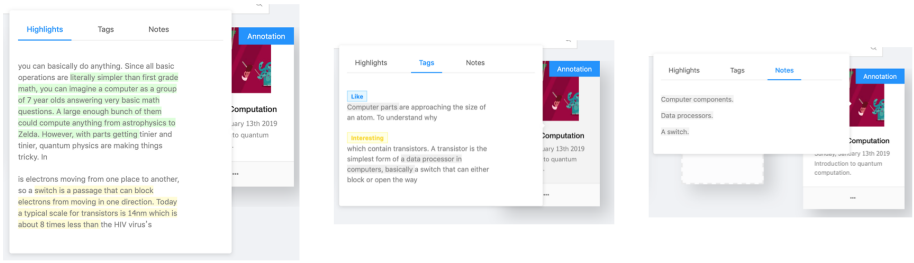


Figure 3.10: Annotation manager.

3.2.3 Annotation Manager

In the course page, each lesson/video is equipped with an annotation manager (AM) which manifests as a blue rectangle at the top right corner of the lesson card, as shown in Figure 3.3. If the learner clicks the blue rectangle, a window will pop up beside the rectangle. This popup window displays all annotations of the video based on category in three separate tabs: “Highlights”, “Tags”, and “Notes”. As Figure 3.10 shows, there is a list of annotations of the same category in each tab.

For each item in the “Highlights” tab, the highlighted transcript segment is the same as in the video player page — a text with a color background; moreover, the preceding and following ten words (if there are) are also displayed to give the highlight a context. If the learner clicks on text with a color background, he/she will be led to the video player page with the video playing from the beginning of the highlighted segment.

In the “Tags” tab, each item consists of a tag sign showing the tag name and the transcript segment that the tag attaches to. The tagged transcript segment is displayed with a grey background and with context, similar to the “Highlights”

tab. The color of the rectangle is the same as the color of the tag icon in the video player page. If the learner clicks on text with a grey background, he/she will be led to the video player page with the video playing from the beginning of the transcript segment that is associated with the tag .

In the "Notes" tab, each note is displayed with a grey background. Clicking on any note will lead the learner to the video player page with the video playing from the beginning of the transcript segment that the note is added to.

3.3 Measures

3.3.1 Task Performance

Task time

Time the subject spent on each searching task was defined as the duration between when the subject opened the AM or clicked on the cover frame of the video, and when the subject was informed of the successful completion of the task by the experimenter.

Annotations

Annotations the subjects made for each video were collected and analyzed.

3.3.2 Surveys

Survey questions and interview questions mainly covered 3 topics: video learning, video annotations, and the AM.

The pre-experiment survey focused on participants' prior experiences with video watching, video learning, and annotating both text and video.

The session survey was used to solicit subjects' immediate reflections on the session, i.e., the video watching process, the annotation process, annotations made, and the experiment material itself (the video).

The post-experiment survey was for the participant to report their experiences in all sessions and general perceptions, including perceptions of the annotation process, annotations, and the AM.

3.3.3 Interviews

The interviews were semi-structured. All subjects were free to express their experiences with the interface or the experiment, but a fixed set of ten questions were asked at some point in the interview. These ten questions focused on how they created their video annotations, how they searched for to-be-located information, and their experiences with the AM.

Chapter 4

Results and Discussion

This chapter presents and discusses data collected on the measures described in the previous chapter. A more integrative discussion of data on these measures is provided in Chapter 5.

4.1 Task Performance

Subjects' performances on the searching tasks were analyzed to investigate whether annotations helped learners locate previously seen information in video, and whether the use of the AM saved time in the process. Time is measured in seconds in this section. A fraction of the user study data can be found in Appendix C. Appendix D shows the R code of data analysis using linear mixed effects models which will be introduced shortly.

4.1.1 Training Session

In the training session, all subjects watched the same video and finished the same set of searching tasks. Thus, individual differences, if any, are supposed to be salient in this session.

All of the 16 subject performances on the 12 tasks of the training session are shown in Figure 4.1. There was not much difference among performances of the 16 subjects, but S7 took a longer time to finish the tasks, giving reason to consider individual differences in the present analysis.

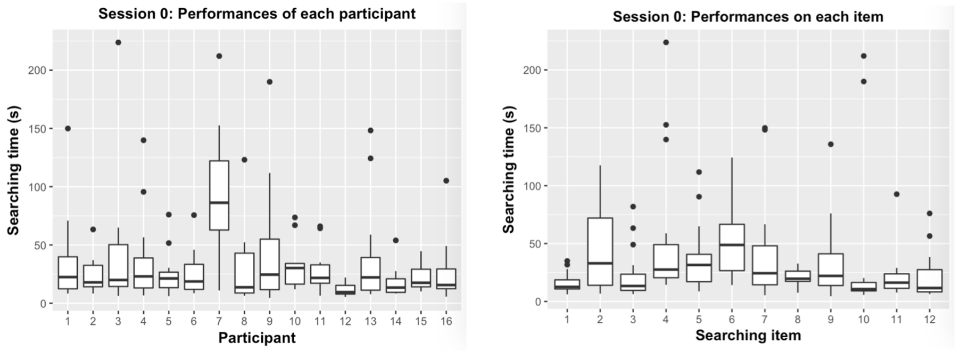


Figure 4.1: Subjects' task performances in the training session.

Subjects performed almost equally on the 12 tasks, with tasks finished using the AM (*mean (*M*)* = 23.43, *standard deviation (*SD*)* = 20.23) took less time than not using the AM ($M = 44.65$, $SD = 47.55$). Locating annotated items ($M = 29.15$, $SD = 33.96$) also took less time than non-annotated items ($M = 42.97$, $SD = 43.21$). Non-annotated items that were also not searched using the AM took the longest time ($M = 53.12$, $SD = 51.69$), while searching for annotated items using the AM took the least amount of time ($M = 21.82$, $SD = 21.25$). Thus, whether the AM was used or not and whether the item was annotated or not might have affected searching time, and there may be an interaction between whether or not the item is annotated and the use of the AM.

To better explore the data, a linear mixed effects model (LMEM) was fitted, as explained at the end of this section. The LMEMs for all sessions are similar, so the one for this session is explained here as a background for following sessions. LMEM was used because it is able to deal with unbalanced designs; in the proposed experiment, the number of annotated items and the number of non-annotated items were unequal, and the number of cases when the AM was used was not the same as when the AM was not used. Thus, LMEM is preferable to traditional analysis of variance (ANOVA) models. Kenward-Roger's approximation was used to obtain p values because it produced acceptable Type I error rates for small samples.

In the LMEM used in this session, between-subject issues were addressed by random effects as they were expected to be generalized. For example, the result should be robust for other participants and for other videos or searching items.

Within-subject issues were addressed by fixed effects, including AM conditions, whether annotated or not, task order, and interactions among those factors.

The analysis results indicated that using the AM significantly reduced searching time (F (1, 9.13) = 5.52, p = 0.043), and searching for annotated items took significantly less time than for non-annotated items (F (1, 130.62) = 5.67, p = 0.019). The effect of the interaction between the AM condition and whether the item was annotated was not significant. The order of the item did not have a significant effect on task performance, nor did its interaction with the AM condition, indicating that the learning effect did not affect subjects' task performances significantly.

Model Explanation

The LMEM used for this session was as follows:

$$time = AM + annotated + AM{:}annotated + task\_order + AM{:}task\_order + (1|subject) + (1|item) + \varepsilon$$

where there were five fixed effects followed by two random effects and an error term on the right side. *AM* was the AM condition associated with the item; *annotated* was whether the item was annotated by the subject while watching the video; *AM:annotated* was the interaction between *AM* and *annotated*, because using the AM for annotated items may be more efficient than for non-annotated items; *task_order* was the order of the task in the set of twelve searching tasks for the video; *AM:task_order* was the interaction between *AM* and *task_order*, because the more the subject used the AM, the more comfortable he/she may feel about using it, potentially resulting in the AM being a more useful tool; $(1 \mid subject)$ was the random intercept for each subject, because the individual difference among subjects may play a role in how they perform on searching tasks; $(1|item)$ was the random intercept for each item, because some items might be easier to find than others; the error term represented the deviations from the predictions due to "random" factors that were out of the purview of the experiment.

4.1.2 Formal Session 1

Subjects' performances in this session are shown in Figure 4.2. As seen in the training session, there was no significant interaction between the AM condition

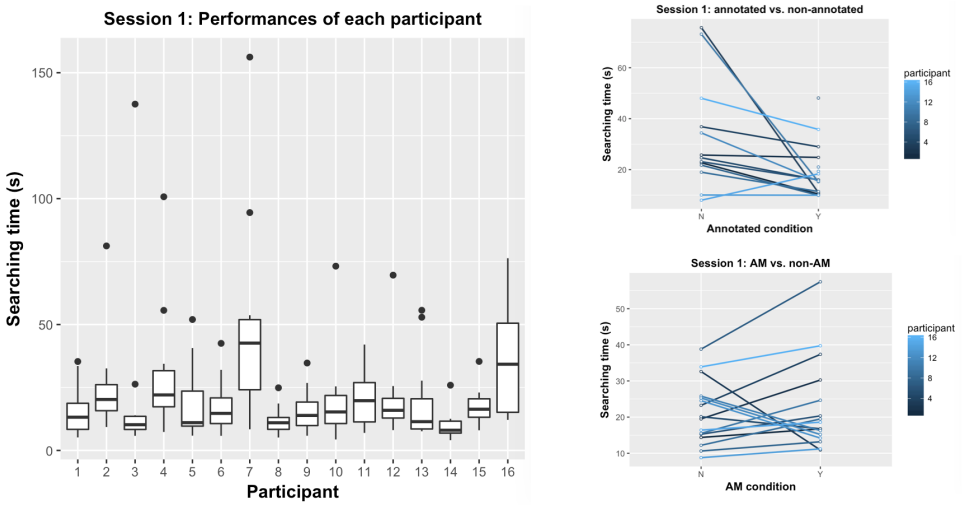


Figure 4.2: Subjects' task performances in formal session 1.

and whether the item was annotated or the order of the item. Whether the item was annotated has a significant effect on searching time (F (1, 169.3) = 5.09, p = 0.025). Searching for annotated items (M = 20.67, SD = 20.03) took significantly less time than for non-annotated items (M = 26.87, SD = 25.12). Using the AM (M = 22.59, SD = 22.61) took about 2 seconds longer to finish the searching task compared with not using the AM (M = 21.07, SD = 19.63), but the effect was not significant.

Model Explanation

The LMEM used for this session was the following:

$$time = AM + annotated + AM{:}annotated + task\_order + AM{:}task\_order + (1|subject) + (1|video/item) + \varepsilon$$

where $(1|video/item)$ represented the random intercept for each item of a video. The factor *video* had 2 levels, because there were two videos in this session; the factor *item* had 12 levels and was nested in the factor *video*.

4.1.3 Formal Session 2

The LMEM used to analyze session 1 data was used for this session. Results suggested that the interaction between using the AM and whether the item was annotated was not significant (F (1, 182.69) = 3.20, p = 0.075). There was no significant difference in searching time between using the AM and not using the AM; using the AM (M = 15.79, SD = 17.39) took almost the same time as not using the AM (M = 15.61, SD = 10.71) to locate information in video. Searching for annotated items (M = 14.79, SD = 16.14) took less time than non-annotated items (M = 19.87, SD = 14.47), and the effect was significant (F (1, 148.63) = 4.69, p = 0.032).

This session was also designed to find subjects' preferences of whether to use the annotation manager to finish searching tasks when there were no AM conditions assigned to tasks. For the training session and session 1, subjects were required to use the AM for half the tasks and not to use the AM for the other half regardless of their preferences and their memories about whether they had annotated related parts or not. So the two sessions were designed to find out how much using the AM could improve subjects' performances on the searching tasks. In session 2, the subject could decide whether to use the AM or not. As shown in Table 4.1, 12 subjects used the AM for more than half of the searching tasks, and the average number of AM cases was 9 while the average number of non-AM cases was 3. This indicates that using the AM was highly preferred over the traditional way of searching for watched video segments by playing the video from the very beginning and sliding through the timeline.

Hypothesis 2, which focuses on the relationship between subjects' preferences of using the AM and annotations they made on the video, is proposed: the more annotations they have made, the more likely they would use the AM to find video segments. As shown in the table counting annotated items and non-annotated items above, 15 subjects annotated more than half of the items in session 2, and the average number of annotated items across 16 subjects was 10 while the number for non-annotated items was 2. Comparing the two rows counting the number of annotated items and items searched using the AM (shown in Figure 4.3), it is not difficult to see that the subject who used the AM the least (1) also annotated the

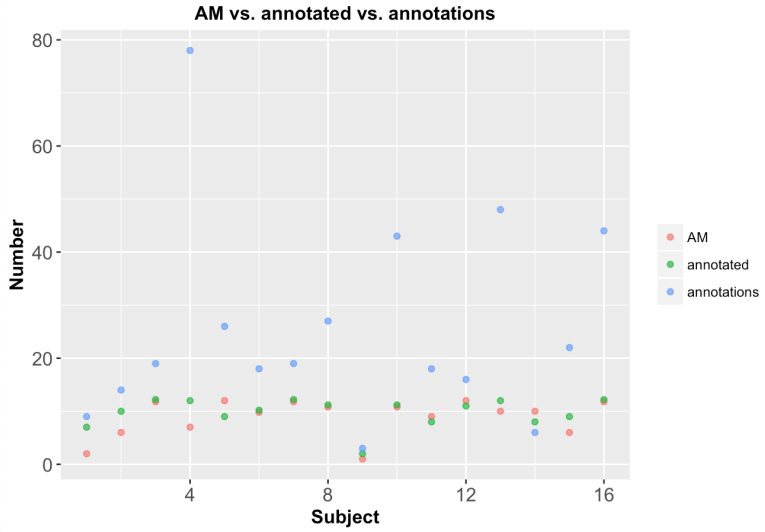


Figure 4.3: Comparing numbers of items searched with the AM, annotated items, and annotations made by subjects.

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	Mean
AM	2	6	12	7	12	10	12	11	1	11	9	12	10	10	6	12	8.93
annotated	7	10	12	12	9	10	12	11	2	11	8	11	12	8	9	12	9.75
annotations	9	14	19	78	26	18	19	27	3	43	18	16	48	6	22	44	25.63

Table 4.1: Numbers of items searched with the AM, annotated items, and annotations by subjects in session 2.

least number of items (2); subjects who used the AM most were also among those who annotated the most. A possible explanation for this phenomenon would be: subjects might remember what and how much they had annotated, thus they may have been able to predict whether they would find helpful information in the AM to help them finish the task, and then made decisions on whether or not to use it based on the prediction or judgement. This hypothesis will be checked in Section 5.1 of Chapter 5 using data from surveys and interviews.

4.2 Annotations

Previous research has shown that learners are not active in making video annotations [19]. This may be due to the lack of a mechanism for later use of these annotations [30]. As the AM supports the use of video annotations for locating

	Highlights	Tags	Notes	Highlight colors	Tag names	Total
Session 0	14.25	2.88	3.94	2.31	1.38	21.06
Session 1	16.56	1.44	4.00	2.44	1.00	22.00
Session 2	20.25	1.44	3.94	2.38	1.19	25.63
Average	17.02	1.92	3.96	2.38	1.19	22.90

Table 4.2: Average numbers of annotations made by 16 subjects.

in-video information, learners' annotating behaviors may also be changed. Thus, annotations that learners created while learning with videos are analyzed here.

As Table 4.2 shows, subjects made a considerable amount of annotations for each video, even though all videos were only around 6 minutes long. On average, subjects made 23 annotations per video, of which 17 were highlights, 2 were tags, and 4 were notes; these proportions of annotations were consistent across all three experiment sessions. Slightly more annotations were made in session 2, and this was a result of the increased number of highlights.

Each subject's preference of annotation types and amount of annotations made were consistent across all three sessions, including the training session. Highlights were considerably more popular than the other types of annotations, which was in line with the previous study about text annotations [49]; 10 subjects made more highlights than tags or notes, 4 subjects made more notes than tags or highlights; no subject made more tags than highlights or notes. The average number of colors selected for highlights was 2 for each of the three sessions, and the average number of tag types used was only 1 for each of the three sessions.

4.3 Surveys

4.3.1 Pre-Experiment Survey

As shown in Appendix A.2, this survey asked about prior experiences with the ViDex platform, video watching, video rewatching, note-taking for videos, and annotating textbooks. Likert-scale questions were the majority, complemented by some questions asking subjects to write down their answers. Not all subjects answered every question. Subject responses to the Likert-scale questions are summarized in Table 4.3.

Question	Scale	Mean	Median	SD	n
Familiarity with ViDex	1-5	1.81	1	1.22	16
Frequency of video watching	1-7	6.18	7	1.25	11
Frequency of video watching for learning	1-7	3.7	4	1.95	10
Frequency of video rewatching for learning	1-5	1.9	1.5	0.99	10
Perceived difficulty of finding previously seen video segments for learning	1-7	1.83	2	0.75	6
Frequency of note taking while watching videos for learning	1-5	2.78	3	1.09	9
Frequency of reviewing notes taken while watching videos for learning	1-5	2.3	3	0.95	10
Frequency of note taking while reading books for learning	1-6	4.3	4.5	1.42	10
Frequency of highlighting while reading books for learning	1-6	3.1	3	1.85	10
Frequency of reviewing annotations (highlights and notes) made while reading books for learning	1-5	2.7	3	0.95	10

Table 4.3: Subjects' responses to pre-experiment survey questions.

Generally, subjects were unfamiliar with the ViDex project. 6 subjects were somewhat familiar with the project, but it is worth noting that the interface used in the presented experiment was new, thus being familiar with the project did not necessarily mean that they were familiar with the interface. Subjects can be seen as novices of the proposed interface in this experiment.

Subjects watched a variety of videos daily, but only watched videos for learning half the week. Not only were course-related videos included, but videos of personal interest, such as DIY tutorials and music videos, were also watched for learning. Most videos were watched on YouTube. Thus, the subjects can be seen as experienced video consumers and intermediate video learners.

Rewatching a video for learning was common but not frequent, and subjects explained that this depended on the difficulty of the video content. 6 subjects provided more detailed reflections on their rewatching habits. Two main purposes for rewatching a video were to prepare for an exam and to review contents that were not understood. Rewatching usually happened the day before an exam, and rewatching both specific segments and the whole video were common. To find a specific segment to rewatch, scrolling through the video was commonly performed, and subtitles as well as thumbnails were thought as helpful in the process.

Writing notes while watching videos for learning was common, especially for parts that were difficult to understand or follow. Those notes, or parts of them, were usually reviewed to consolidate memory or to prepare for an exam or assignments.

In comparison, writing notes while reading books was performed more frequently, and notes were usually written on main points or information that was not already known. Highlights were made less than notes for books, and part of the reason for this was because highlighting would reduce the book's resale value. Though highlighting helped to bring attention to specific parts of the text so that the learner did not need to read or review all of the text, highlighting was considered as more passive than writing notes. Subjects reviewed either all of their annotations or part of their annotations on books slightly more often than notes on videos for exams or assignments.

In summary, responses to this survey suggested that (1) subjects could be considered as novice users of the proposed interface in this experiment, (2) subjects were experienced in video watching and had some experience with video learning, (3) subjects usually used the low-level strategy of scrolling through the video timeline to find specific video segments for revisitation, (4) most subjects annotated for both videos and books, while more subjects reviewed annotations on books than notes for videos. Discussions of annotating video vs. text and how the proposed interface affected subjects' video reviewing patterns will be given in Section 5.1 of Chapter 5, with evidence collected from interviews and other surveys.

4.3.2 Session Survey

Session surveys filled right after subjects finished watching the video solicited immediate reflections on learning experiences, as shown in Appendix A.3. Subject responses are summarized in Table 4.4. All questions of this survey were Likert questions with a 1-5 scale.

Generally, subjects had some, but not ample, prior knowledge about the topic covered in the video (M = 2.88, SD = 1.36), which simulates the situation in the classroom when an instructor gives a lecture. Subjects were interested in the videos (M = 3.91, SD = 0.92) and topics they covered (M = 3.95, SD = 0.79), and the videos were thought to be well-produced and effective in teaching the topics (M =

Question	Mean	Median	SD	n
Prior knowledge about the topic.	2.88	3	1.36	16
Interest in the topic	3.95	4	0.79	11
Interest in the video	3.91	4	0.92	11
Perceived effectiveness of the video in teaching the topic	4.18	4	0.8	11
Focused more on learning than on making annotations	3.5	3.5	0.95	16
Learning took more efforts than making annotations	2.64	2	1	11
Perceived helpfulness of making annotations in learning	3.81	4	0.82	16
Perceived quality of learning	3.69	4	0.69	16
Perceived familiarity of the video content after learning	3.64	5	0.85	11
Perceived possibility of reviewing annotations made in the session for exams of the course	3.97	4	1.06	16

Table 4.4: Subjects' responses to session survey questions (Likert scale, 1 to 5) of session 1 and 2.

4.18, $SD = 0.8$).

In terms of annotating while learning, subjects focused more on learning than annotating ($M = 3.5$, $SD = 0.95$), but this depended on how difficult the video content was. The more difficult the video was, the more efforts were devoted to learning. Considering that the three videos used in the experiment were rated at medium to high level of difficulty in early pilot studies, and that subjects generally did not have much prior knowledge about the topics covered by those videos, the annotation process should not have significantly diverted the learner's attention from learning the video content. Generally, subjects thought the annotation process was helpful for learning ($M = 3.81$, $SD = 0.82$), and would like to review their annotations for potential exams of the course ($M = 3.97$, $SD = 1.06$). Thus, subjects may be willing to see their annotations again either in the AM or on the video player page.

4.3.3 Post-Experiment Survey

The post-experiment survey used in the presented experiment can be found in Appendix A.4. Subjects' responses to the post-experiment survey are shown in Table

Question	Mean	Median	SD	n
Ease of use of the AM	4.19	4	0.54	16
Like using the AM	4.19	4	0.75	16
Using the AM saves searching time	4.44	5	0.73	16
Will use the AM to rewatch video segments for learning	4.19	5	1.22	16
Making annotations while watching videos for learning is easy	3.94	4	0.68	16
Getting used to making annotations while watching videos for learning is easy	4.3	4.5	0.95	10
Will review annotations made on videos for academic goals such as to pass exams	4.38	4.5	0.81	16
Annotations were made more for learning than for the experiment tasks	3.56	4	1.09	16
Will make annotations for course videos in a similar way as in the experiment	4.45	4	0.52	11

Table 4.5: Subjects' responses to post-experiment survey questions (Likert scale, 1 to 5).

4.5. This set of Likert questions used a scale of 1-5.

The AM was considered easy to use and effective in saving searching time. The AM was also liked by subjects, and was projected to be used for future rewatching. Generally, the annotation process was also thought of as easy to become familiar with.

During the annotation process, most subjects (13 out of 16) were making more annotations for learning than for the tasks, and they felt strongly that they would make annotations on the proposed interface the same way for learning with course videos. Subjects also reported the strong preference to review annotations they had made for academic goals, such as passing an exam.

4.4 Interviews

Interviews were carried out to have a better understanding of subjects' annotating patterns, perceptions of different types of video annotations and the annotation process, video rewatching patterns, and usage patterns and perceptions of the AM.

	Positive	Negative	Neutral
Highlights	10	2	0
Tags	2	4	0
Notes	5	6	1
The annotation process	7	6	1

Table 4.6: Summary of subjects' interview responses regarding annotations and the annotation process.

4.4.1 Annotations

Two topics were discussed by subjects: the content of their annotations, and their perceptions of the three types of video annotations. Subjects' responses are summarized in Table 4.6.

Regarding the content of annotations, 13 out of 16 subjects said that they mainly annotated on key points or important points of the video. Subjects also reported that they had annotated on more detailed information such as examples, definitions, or descriptions (n=4); one subject annotated on things that she liked; two subjects annotated on words that were not understood.

Subjects had mixed opinions toward the three types of annotations.

For highlights, advocates (n=10) thought they were easy and quick to make while watching the video and they emphasized more on the original material so that learners knew what information they were annotating on. Highlights were also thought to be useful to label or bookmark main ideas of the video, and they reminded the learner of information following or preceding them as well, which helped in recalling and finding information in the video. For other subjects (n=2), highlights were random and not useful, by being too broad, not specific enough, and unable to stimulate thinking.

Tags did not receive as much positive feedback as highlights. One subject referred them as interesting, but did not think they would be helpful. One subject referred to tags as more personalized than highlights but not as useful as notes in getting information, and this comment is in line with previous studies comparing tagging and handwritten notes [67]. Another positive feedback came from a subject who thought that tags were helpful in providing the context of an annotation, just as highlights were. In addition to being considered unhelpful and cumbersome,

slowing down the video watching process, tags also caused some confusion. S12 commented on tags as follows:

> "Tags will be a bit harder, because it's hard to categorize it. So if I want to flag it as important, I would have to probably create a new tag, (and then define) what type of 'important' it is."

Similar to notes, tags are learners' comments on the original material. Even though learners add their own input by adding either notes or tags, the difference between these two types of annotations is that a tag is reusable once created. Tags are usually shorter than notes as well and are usually words, while notes are usually full sentences.

Subjects' opinions on notes varied. In the opinion of supporters (n=5), notes enabled them to reinterpret information in a way that made more sense to them and made the information more straightforward; notes also helped them organize information, such as grouping separated information, so that they did not need to make two or more highlights; notes were used as precursors or place marks as well. Opposite opinions (n=6) focused on the process of creating notes and the contents of notes. The subject who referred to tag-making as a cumbersome process also made the same comment for notes. Another subject preferred handwritten notes over taking notes on computers. Regarding the contents of notes, two subjects felt that there was no need to add notes because all information was already in the transcript, and one of them would rather add notes from other sources found online for future reference if necessary; one subject thought that notes were "other things" added to the original material, and thus preferred highlights; one subject made mainly notes and highlights because notes alone could not remind the subject what they referred to.

4.4.2 The Annotation Process

In this part of the interview, subjects were asked about their perceptions of the process of making annotations while watching a video for learning. 14 subjects answered this set of questions.

Half of the subjects thought making video annotations facilitated learning (n=7). Annotating on videos was thought to be natural and easy instead of being a disrup-

tive process. The interaction with the text and the video also made the video watching process more engaging rather than a passive information receiving process, and annotation "footprints" of this process helped to find segments for future rewatching, like pinpoints. Subjects also reported that annotations helped them remember information in a more clear and organized way. For example, S15 commented as follows:

> "I think it helped, like, making sure that I understood the information, because then I would go back to the point that I thought was important, and reread it. So it's like double the information."

The annotation process was also reported to have an effect on how subjects found information in video, and S6 commented as follows:

> "The behaviour of annotating, highlighting the key points or concepts build that road map in your brain. So when you have to go back and find certain things, the active highlighting that builds that structure in your head so you know roughly where to look."

S8 also reflected that she was well aware of what annotations to make by going back and forth, and structuring her annotations, thus she was able to recall whether she had annotated the to-be-located information or not.

The annotation process for video watching was also thought to lead to active learning because subjects needed to make annotations physically, and this process as well as annotations themselves helped in remembering video content and keeping up with the progression of the video (n=5). S8 reflected as follows:

> "In a video, he (the speaker) goes through things in a certain order. He talks about certain things. If I don't make notes, then when I am trying to remember what was going on, I can't remember the sequence of the topics, I can't remember exactly what details he described, and I can't remember the exact location of, for example, the function of hypothalamus."

However, the multi-sensory and transient nature as well as the highly interactive process posed challenges to subjects in focusing on the content. 3 subjects said

that they would prefer watching the video for the first time without doing anything. Then they would go back and make annotations during their second-time viewing, and then a third viewing without doing anything again. The 3 presentations of the information, the transcript, the filmstrip, and the video, competed for subjects' attentions, thus it was difficult for some subjects to focus on all of them. The intrinsic pace of a video and subjects' pace of making annotations posed an added difficulty as well. For example, S12 commented as follows:

> "Well, while I was making annotations, I was focusing on writing it, so at that point I am just hearing things, or focusing on what I'm writing, or like the dialogue outside, as opposed to focusing on what's on the screen. So with the case of the first, there was a lot more information displayed on the screen, I couldn't focus on it while I was writing annotations, but I could still hear everything that is being said.
>
> .....
>
> It's extremely hard to take notes and pay full attention to the video at the same time. If it was just audio, that would be different, because you still hear everything. But if you're writing things down, you kind of fall behind at some point, because you can't write faster than they speak."

When unbalanced attention was paid to the three presentations of information, the video watching experience would be completely changed. For example, S11 commented as follows:

> "Right now I know I wasn't paying attention to the visuals, I was just looking at the text, so I was just like reading a book than really watching a video. That's what it felt like to me. "

S14 commented on his annotation style as "out of an instinct" and there was no specific plan on how to annotate. He also reflected that he played back frequently as a result of a disrupted stream of thoughts caused by the annotation process.

	Support	Not Support
Annotating text was easier	5	3
Annotating video was more engaging	5	3
The AM was helpful for navigating to video segments of interest	14	0
Annotations affected the use of the AM	8	–

Table 4.7: Numbers of subjects' interview responses regarding annotating video vs. annotating text and the use of the AM.

4.4.3 Annotating Video vs. Annotating Text

Subjects were asked about their perceptions of the difference between the annotation processes for video and text. 12 subjects answered this questioning, and a summary is given in Table 4.7.

Subjects' opinions were mixed for either type of annotating. For some subjects (n=5), annotating text was easier and less disruptive than annotating videos, and they were able to read and annotate at their own pace while annotating text rather than trying to keep up with the progression of the video. While annotating text, subjects also needed to pause reading, annotate, and resume reading, just as they needed to pause watching, annotate, and resume watching while annotating video, but pausing and resuming text reading was perceived as more comfortable because subjects did not need to slow down or play back, so the whole reading and annotating process was more integrated. However, S3 thought about it in the opposite way: annotating while reading took more time as it was impossible to read and annotate on text at the same time. Several subjects (n=3) wrote notes in a separate notebook while reading books, thus it took more time for them to link their notes back to the original text segments to review; they had to search through the book more intensely as they could not skim through the book like skimming through a video in the experiment. For example, S13 commented as follows:

> "When I read, I usually take notes on a separate piece of paper, so it's not on the book itself, which takes longer to match my notes to the section or the paragraph or whatever."

Except for providing quicker links from annotations back to the original video segments or transcript segments, making video annotations was also thought of as a more engaging learning process than annotating text because of the added audio

and visuals (n=5). S15 commented as follows:

> "I like the visual component of having a video while also making annotations, rather than just reading the same words, then reread it a bunch of times but with only the annotations."

Though some subjects thought annotating, especially highlighting, videos was quick and easy, 3 subjects regarded this process as more disruptive and distracting than annotating on text. More specifically, frequent pausing and playing back the video disrupted subjects' flow of video watching and learning.

4.4.4 The Use of The AM

This section of interview questions asked subjects about their strategies using the AM in session 2 when subjects made their own decisions regarding whether to use or not use the AM. General perceptions of the AM were also asked. 14 subjects talked about their use of the AM in the interview. A summary is given in Table 4.7.

All subjects (n=14) thought that the AM was helpful because it took them directly to video segments of interest so that they did not need to scan through the whole video again, including the two subjects who used the AM only once and twice respectively in session 2.

Annotations and the annotation process seem to be closely associated with the use of the AM, and the quality and amount of annotations were reported to impact the effectiveness of the AM. Subjects mainly annotated on key points or concepts of the video content, and they were aware of what information they had annotated on (n=7), thus looking at annotations gave them clues about the structure of the video content, and also the location of the to-be-located information if it was not already in the AM. For example, S12 reflected on the process of finishing tasks in session 2 as follows:

> "I wanted to find just the four main parts of the brain, so I annotated it, you know. I went over brainstem, thalamus, cerebellum, and cerebrum. Those were like the four main topics in the video, right, which made it way easier to jump. Everything that could be said can be classified under these four. So you already know which quarter it is gonna be in. I feel like most of my times were pretty fast in the last one."

The AM was used extensively even when subjects were not confident about their familiarity of the video content. For example, S16 annotated on parts that were not understood, and she went to the AM everytime when presented with a searching item that she did not understand. So, the AM was used like a placeholder for certain types of information. When subjects did not know much about the topic, they generally could not identify which parts were important, thus they tended to annotate non-understood parts; however, annotating too frequently on non-understood parts could lead to ineffectiveness of the AM. S15 reported being "swamped in annotations" for a video that she did not know much about.

Half of the subjects (n=8) did not use the AM to search for pictures, but S8 used it for pictures because her annotations reminded her of the visuals of the video. S6 only used the AM for the key points and did not use the AM for extra details, which the subject referred to as "filled-in stuff".

Regarding future use of the AM, most subjects thought that it would be useful in directing them to video segments and reminding them of the structure and contents of the video (n=9). S12 thought that ideally he would not need to revisit video segments if he had made good enough annotations. S7 imagined that she would use the AM to rewatch video segments if asked questions that she was unsure about. Both S6 and S9 thought that the AM would be used more often if they were unfamiliar with the topic of the video, because the AM provided convenient access to parts of the video that they were struggling with. S16 would like to see screenshots of video frames along with text in the AM.

4.5 Summary

To answer the research question and understand related issues, both quantitative data and qualitative data were collected and analyzed. Quantitative data includes time spent on tasks and responses to most of the survey questions. Qualitative data consists of responses to interview questions and some survey questions. These different types of data complement each other and jointly answer the research question in a richer way.

Each session of the experiment and corresponding data collected serve for their own purposes. The training session was used for the subject to develop their own

strategy, and the data analysis for this session was used to reveal possible interesting patterns. Both the formal session 1 and 2 were used to compare the AM condition vs. the non-AM condition, and to compare searching for annotated information vs. non-annotated information. Results suggested that the use of the AM did not have a significant effect on searching time, but searching for annotated information took significantly less time than for non-annotated information. The formal session 2 was also used to find out whether subjects would use the AM spontaneously, and results indicated that the AM was highly preferred and used frequently by subjects; in addition, the number of annotations made by a subject might lead to the frequent use of the AM, which will be discussed further in the following chapter.

Subjects' responses to survey questions and interview questions provided profound data on video annotations, the annotation process, and the AM. The pre-experiment survey outlined the profiles of subjects and their habits of rewatching videos and reviewing annotations for both books and videos. Session surveys solicited prior knowledge of the topic and perceptions of the video, and immediate reflections on the annotation process and video annotations made for the video. Responses to post-experiment survey questions indicated that the AM was rated high in usability and usefulness, and was projected to be used in the future. Data from subjects' responses to interview questions revealed that opinions on the three types of annotations and the annotation process were mixed. Interview data also suggested that both the annotation process and video annotations were closely related to how subjects located in-video information.

Chapter 5

General Discussion

To answer the research question, a novel interface has been designed and implemented, based on which both quantitative data and qualitative data have also been collected and analyzed. Main findings are summarized as follows:

1. Video annotations seem to be effective in helping learners find previously seen video segments;
2. The annotation manager is highly preferred by learners over the traditional way of locating information in video;
3. Integrating text with video and the annotation process based on integration have the potential to improve both the video learning experience and learning outcomes.

The proposed interface is novel as it presents user-created annotations of multiples videos on the same page, the course page, and provides quick accesses to annotated video segments.

The rest of this chapter checks the two hypotheses and discusses results from task performances, survey questions, and interview questions more integratively. These discussions offer richer explanations of the results, because session surveys, which were filled right after subjects finished learning with videos, recorded subjects' immediate reflections on their learning and annotating experiences. These responses will partially explain subjects' task performances; responses to the post-experiment survey and interview questions solicited general perceptions of the

proposed interface and the experiment, and will provide valuable data on issues related to the research question, while also partially explaining subjects' task performances.

5.1 Interpreting the Results

Annotations Helped in Locating Information

Searching for annotated information was significantly faster than for non-annotated information in all of the three experiment sessions. So how is it that annotated information was located faster? A closer look at the searching process may give an explanation. To search for a given piece of information, the subject started from the course page and entered the video player page to locate the information either through the AM or by clicking on the cover frame of the video. In the case of using the AM, the subject started with annotations he/she had made on the video, and then entered the video player page; in the case of not using the AM, the subject entered the video player page directly. The subject may have been able to locate the information right after entering the video player page if he/she found the corresponding annotations in the AM, otherwise he/she needed to have searched for the information on the video player page. On the video player page there were indications of annotations in all of the three sections, thus the subject was reminded of where annotations were and what had been annotated on, no matter which section they focused on in the searching process. All subjects made a considerable amount of annotations mainly on key points, which were also what most of the searching items were about, thus annotations created saliency of key points, which might have made annotated information easier to locate.

According to Guthrie's cognitive model, locating information in text requires 5 steps: (1) formulation of a goal, (2) inspection of appropriate categories of information, (3) planning the inspections of information, (4) extraction of details, and (5) recycling to obtain information [29]. Subjects made significantly more highlights than tags or notes, and highlights were integrated with the original material better than the other types of annotations, thus indications of annotations on the player page may have supported subjects in step (2) and step (4), which made the

searching process smoother and quicker. The results are also in line with those from Salome's experiment in which the table of contents of the video and indications of chapters on the video timeline were found to be helpful in locating information in the video [13, 59]. Moreover, the process of making annotations stimulated subjects to think about the video contents and process the video contents at a deeper level, thus reinforced memory and accelerated the searching process. This was revealed by subjects' reflections on the annotation process. Thus, Hypothesis 1 is validated.

The AM Was Highly Preferred

The AM was highly preferred, but only data from the training session revealed its effectiveness in reducing searching time. In the training session, which was used for subjects to become familiar and develop their learning strategies with the interface, the AM conditions assigned to tasks were the same for all subjects. So, it is possible that some tasks were inherently easier to be located with the AM. In the formal session 1, each half of the subjects watched one of the two videos, and the AM conditions assigned to tasks of each video were counterbalanced, which meant that 8 subjects watched the same video, but only 4 of the 8 subjects finished searching tasks with the same AM conditions and the other 4 subjects' searching tasks were with the reversed AM conditions, i.e., a task was assigned with the AM if it was not assigned for the other half of the 8 subjects who watched the same video.

The case was different for the formal session 2, in which subjects used all features at will. Subjects took a longer time to finish searching tasks when using the AM, though the result was not significant. The subjects' motivations of using the AM in the formal session 2 may shed light on this. It is possible that subjects relied on the AM as an extension of their internal mental model of the video content, and as a result had difficulty finding information without the help of the AM on the video player page [13]. For example, S16 knew that she annotated on parts that she did not understand, so she went to the AM every time when needing to search for a piece of information that she did not understand. It is also possible that the searching process would take even longer if the AM was not used. Results from

pre-experiment surveys suggested that subjects relied on low-level strategies such as scrolling the timeline to find a specific video segment, however, the spontaneous use of the AM reflected subjects' needs to process video information at a macro-level instead of solely at the micro-level. This finding is in line with a number of previous studies [43, 53, 64].

Annotations and the Use of AM

Hypothesis 2 highlights the effect of annotations on the use of the AM. Results from session 2 revealed that subjects who annotated the least number of items were also among those who used the AM the least. A possible explanation for this phenomenon would be: subjects might remember what and how much they had annotated, and thus had some predictions about whether they would find helpful information in the AM for finishing the task, and then made decisions on whether to use it or not based on their prediction or judgement. Subjects' responses to related interview questions (summarized in Section 4.4.4) provided evidence for this explanation. Subjects did remember what they had annotated, and the process they went through to make these annotations built a "roadmap" of the video content in their mind. Looking at annotations in the AM reminded subjects the structure of the video content and locations of the to-be-located information in the original video. Therefore, Hypothesis 2 is confirmed.

The Annotation Process

The annotation process, as a type of macro-level processing which enables the learner to temporarily step away from the current point of the video being played and to look at the video from a more global and personal perspective, has the potential to improve video learning experience and learning outcomes by overcoming the intrinsic challenges of learning with videos. This has been shown in subjects' reflections of their learning experiences in the experiment. Though the incorporation of the annotation process into video learning may pose challenges to the continuity of the video viewing process, the added interactivity gives more learner-control, and thus has the potential to help learners with low working memory capacity to better assimilate the video contents [25, 27, 31, 40]. The annotation process also

creates a more engaging video learning environment where learners focus more intensely on learning and can more easily avoid distractions, which can promote self-regulated learning [10].

The Integration of Text and Video

The integration of text in the video-based learning environment has also probably facilitated both the learning process and the in-video information-seeking process. Learners have been found to both read texts and watch videos non-linearly, where parts of the text or the video are visited repeatedly [33, 65]. This calls for a mechanism to support this revisitation. It is easy to locate information while reading text non-linearly, as the entirety of the text, as well as devices such as tables of contents or indexes, are exposed to the reader. However, locating video segments while watching poses more challenges if the viewer is not supported by more advanced functions than the traditional VCR controls. For example, while watching a video, if the viewer wants to rewatch a segment, he/she will need to slide through the timeline of the video to find the segment, while it is easier and faster to locate a previously read paragraph while reading a book. In a video-based learning environment, a timeline and filmstrip can provide continuous progression, whereas a transcript provides discrete progression enabling learners to jump to video segments easily, thus navigate the video more conveniently. For videos that have limited text, it is possible that less highlights will be made, and more notes or tags may be made by learners to add their own inputs as text. In this case, locating information on the video player page will require more efforts, as the learner will have to check what the notes or tags are by clicking on icons. However, the use of the proposed AM should still be able to help in the information-seeking process and provides quick access to specific video segments. Therefore, incorporating various annotation features is critical to accommodate different types of learning, as suggested by previous study [67].

5.2 Limitations

Though the proposed interface design and experiment answered the research question, there are some intrinsic limitations for both of them.

There are four main limitations regarding the experiment materials. First of all, the videos subjects watched were relatively short, being only 5-7 minutes long. This was due to the time constraints of the experiment. In practice, videos for online learning are usually longer [28]. Videos chosen for the proposed experiment were also intense in information covered, which may have posed extra pressure on subjects. Second, the presented experiment used specific types of video containing intense information both visually and textually, so it is uncertain whether the experiment results can be generalized to other videos that are, for example, solely visually rich or textually rich. Third, the subjects were asked to locate specific pieces of information rather than information they wanted to find; while this design makes sure that subjects were exposed to the same tasks, it may still have brought in irrelevant factors which could have deteriorated the results. For example, some subjects might have difficulty understanding the items, and may have spent a lot of time wandering in the AM or on the video player page as a result, rather than putting efforts toward locating the information. Finally, though the number of subjects recruited was in line with similar research studies [2], and the data from the 16 subjects participated in the experiment gave meaningful results, having more subjects will potentially give more generalizable results.

Three limitations are about the methodology of the presented experiment. First, subjects were asked to finish the searching tasks right after finishing learning with the video, whereas in practice learners revisit materials for a longer period of time after their first visit [6]. Thus, longitudinal studies or delayed tests may bring better results [9]. Second, as reading on paper has been found to be preferred over on screen, despite all of the advanced interactions and functions provided by electronic readers [57], the annotations subjects made on the proposed interface may have been more for the text than for the video. Third, though the presented controlled lab experiment ruled out confounding factors, such as distractions from surrounding environments, the experiment is more artificial than deploying the learning system in classrooms or Amazon Mechanical Turk. As a result, subjects may have performed tasks unnaturally in the experiment, potentially not reflecting real life.

Another limitation is on the interface. In the proposed interface, the window of the AM is not of a fixed size, which was reported by some subjects as trouble-

some because they had to scroll to the bottom of the website to examine all their annotations if there were too many. This inconvenience may have deteriorated the efficiency of the AM in helping learners locate in-video information, and a more structured organization of annotations on each tab of the AM window could potentially improve user experience with the interface.

Chapter 6

Conclusion and Future Work

The presented work explores the use of video annotations in in-video information-seeking. To answer the research question, a novel video-based learning and annotating environment has been developed to enable learners to learn with and annotate on both the video and the text, by integrating the visual and the textual material of the video; a novel interface of the annotation manager was also designed to collect annotations that the learner has made on a video and to provide quick accesses to annotated video segments.

A controlled lab experiment with 16 undergraduate students as subjects has been carried out. Experiment results showed that video annotations were effective in reducing time spent on locating previously-seen information in video, and the novel design of the annotation manager was highly preferred by subjects and was used spontaneously and frequently in the information-seeking process. The proposed interface design of integrating text with video was perceived as helpful in video-based learning, and has the potential to improve both the learning experience and learning outcomes. Another contribution of the presented work is the exploration and analysis of different types of video annotations.

Future Work

Though the proposed interface design and experiment design were able to answer the research question and provide meaningful data, a number of future investiga-

tions are needed to further explore video annotations and their uses.

The annotation manager, which enables the learner to see all annotations they have made on a video and provides quick accesses to video segments, was only provided on the course page. Thus, the learner was not able to use it while learning on the video player page. Providing an annotation manager on the video player page has the potential to facilitate information organization and management in the process of learning, which can be a topic for future exploration.

In the proposed annotation manager, there were only pieces of textual information, so incorporating pictures such as video frames can potentially improve its functionality and further improve the annotating experience.

Regarding the annotation process, it is worthwhile to explore a faster and simpler way of creating annotations, so that learners do not need to go through multiple steps to add an annotation. Though having the AM encouraged learners to add more annotations for later use, having a less artificial procedure to add annotations can potentially further encourage learners to make more annotations, as they would not need to divert their attention away from understanding video content.

Subjects in the presented experiment exhibited various patterns of video learning, and it will be valuable to identify types of video learners, in a similar way to identifying types of readers who exhibit distinct reading patterns [33]. Moreover, reading proficiency has been found to have a significant effect on information assimilated from both video and text [24], so it will also be interesting to explore "viewing proficiency" and how this may be related to video-based learning.

Though subjects suggested that the annotation process took more efforts than learning the video content, they also perceived the annotation process as helpful for their learning. Thus, it will be beneficial to investigate the role of the annotation process in video-based learning, for example, by investigating what portions of learners' attention are devoted to the annotation process and how this affects learning experiences and learning outcomes.

Finally, the integration of video and text in the proposed design consists of the video and its literal transcript, which is structured more loosely than an article or a book. Therefore, exploring how the video transcript can be more effective and better utilized for learning will be beneficial to further unlock the potential of a video's transcript and annotations.

Bibliography

[1] M. J. Adler and C. Van Doren. *How to read a book: The classic guide to intelligent reading*. Simon and Schuster, 2014. → page 11

[2] A. Al Hajri. *Shaping video experiences with new interface affordances*. PhD thesis, University of British Columbia, 2014. → pages 22, 59

[3] I. E. Allen and J. Seaman. *Online Report Card: Tracking Online Education in the United States.* ERIC, 2016. → page 1

[4] T. Anderson and D. R. Garrison. Learning in a networked world: New roles and responsibilties. In *Distance Learners in Higher Education: Institutional responses for quality outcomes. Madison, Wi.: Atwood*. 1998. → page 1

[5] O. Aubert, Y. Prié, and C. Canellas. Leveraging video annotations in video-based e-learning. *arXiv preprint arXiv:1404.4607*, 2014. → pages xi, 13

[6] C. Bazerman. Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written communication*, 2(1):3–23, 1985. → page 59

[7] L. Beardsley, D. Cogan-Drew, and F. Olivero. Videopaper: Bridging research and practice for pre-service and experienced teachers. *Video research in the learning sciences*, pages 479–493, 2007. → pages 15, 17

[8] F. Bentley and J. Murray. Understanding video rewatching experiences. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, pages 69–75. ACM, 2016. → pages 3, 7

[9] L. W. Brooks, D. F. Dansereau, J. E. Spurlin, and C. D. Holley. Effects of headings on text processing. *Journal of Educational Psychology*, 75(2):292, 1983. → page 59

[10] K. S. Bull, P. Shuler, R. Overton, S. Kimball, C. Boykin, and J. Griffin. Processes for developing scaffolding in a computer mediated learning environment. 1999. → page 58

[11] P.-S. Chiu, H.-C. Chen, Y.-M. Huang, C.-J. Liu, M.-C. Liu, and M.-H. Shen. A video annotation learning approach to improve the effects of video learning. *Innovations in Education and Teaching International*, 55(4): 459–469, 2018. → page 13

[12] S. Cojean and E. Jamet. Facilitating information-seeking activity in instructional videos: The combined effects of micro-and macroscaffolding. *Computers in Human Behavior*, 74:294–302, 2017. → pages xi, 9, 10

[13] S. Cojean and E. Jamet. The role of scaffolding in improving information seeking in videos. *Journal of Computer Assisted Learning*, 34(6):960–969, 2018. → pages 9, 56

[14] J. Costley, M. Fanguy, M. Baldwin, C. Lange, and S. Han. The role of motivation in the use of lecture behaviors in the online classroom. *Journal of Information Technology Education: Research*, 17(1):471–484, 2018. → pages 3, 7

[15] J. Davis and D. Huttenlocher. Conote annotation homepage. *Available on the World*, 1994. → page 12

[16] S. I. De Freitas, J. Morgan, and D. Gibson. Will moocs transform learning and teaching in higher education? engagement and course retention in online learning provision. *British Journal of Educational Technology*, 46(3): 455–471, 2015. → page 1

[17] E. Delen, J. Liew, and V. Willson. Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments. *Computers & Education*, 78:312–320, 2014. → page 10

[18] S. Dodson, L. Freund, D. Yoon, M. Fong, R. Kopak, and S. Fels. Video-based consensus annotations for learning: A feasibility study. *Proceedings of the Association for Information Science and Technology*, 55 (1):792–793, 2018. → page 19

[19] S. Dodson, I. Roll, M. Fong, D. Yoon, N. M. Harandi, and S. Fels. An active viewing framework for video-based learning. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, page 24. ACM, 2018. → page 40

[20] S. Dodson, I. Roll, M. Fong, D. Yoon, N. M. Harandi, and S. Fels. Active viewing: A study of video highlighting in the classroom. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*, pages 237–240. ACM, 2018. → page 19

[21] B. R. Dye. Reliability of pre-service teachers coding of teaching videos using video-annotation tools. 2007. → pages 15, 18

[22] J. D. Finn and K. S. Zimmer. Student engagement: What is it? why does it matter? In *Handbook of research on student engagement*, pages 97–131. Springer, 2012. → page 1

[23] M. Fong, S. Dodson, X. Zhang, I. Roll, and S. Fels. Videx: A platform for personalizing educational videos. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, pages 331–332. ACM, 2018. → pages v, 28

[24] A. Furnham, S. De Siena, and B. Gunter. Children's and adults' recall of children's news stories in both print and audio-visual presentation modalities. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 16(2):191–210, 2002. → page 62

[25] R. Garner. Strategies for reading and studying expository text. *Educational Psychologist*, 22(3-4):299–312, 1987. → page 57

[26] A. Girgensohn, F. Shipman, J. Adcock, M. Cooper, and L. Wilcox. Locating information in video by browsing and searching. In *Interactive Video*, pages 207–224. Springer, 2006. → pages xi, 8, 9

[27] S. M. Glynn and F. J. Di Vesta. Control of prose processing via instructional and typographical cues. *Journal of Educational Psychology*, 71(5):595, 1979. → page 57

[28] P. J. Guo, J. Kim, and R. Rubin. How video production affects student engagement: an empirical study of mooc videos. In *Proceedings of the first ACM conference on Learning@ scale conference*, pages 41–50. ACM, 2014. → page 59

[29] J. T. Guthrie and P. Mosenthal. Literacy as multidimensional: Locating information and reading comprehension. *Educational Psychologist*, 22(3-4): 279–297, 1987. → page 55

[30] N. M. Harandi, F. Agharebparast, L. Linares, S. Dodson, I. Roll, M. Fong, D. Yoon, and S. Fels. Student video-usage in introductory engineering courses. *Proceedings of the Canadian Engineering Education Association (CEEA)*, 2018. → pages 19, 40

[31] B. S. Hasler, B. Kersten, and J. Sweller. Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 21(6): 713–729, 2007. → page 57

[32] B. Hosack. Videoant: Extending online video annotation beyond content delivery. *TechTrends*, 54(3):45–49, 2010. → page 14

[33] J. Hyönä and A.-M. Nurminen. Do adult readers know how they read? evidence from eye movement patterns and verbal reports. *British Journal of Psychology*, 97(1):31–50, 2006. → pages 58, 62

[34] H. Jung, H. V. Shin, and J. Kim. Dynamicslide: Exploring the design space of reference-based interaction techniques for slide-based lecture videos. In *Proceedings of the 2018 Workshop on Multimedia for Accessible Human Computer Interface*, pages 33–41. ACM, 2018. → pages xii, 18, 19

[35] M. Ketterl, R. Mertens, and O. Vornberger. Vorlesungsaufzeichnungen 2.0. In *Lernen–Organisation–Gesellschaft. eCampus-Symposium der Osnabrücker Hochschulen*, pages 2–5, 2008. → page 15

[36] J. Kim. Toolscape: enhancing the learning experience of how-to videos. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2707–2712. ACM, 2013. → pages xi, 13, 14

[37] M. Kipp. Anvil-a generic annotation tool for multimodal dialogue. In *Seventh European Conference on Speech Communication and Technology*, 2001. → pages xi, 15, 16, 18

[38] D. LaLiberte. Hypernews homepage, 1997. → page 12

[39] R. Lowe. Interrogation of a dynamic visualization during learning. *Learning and instruction*, 14(3):257–274, 2004. → page 8

[40] D. L. Lusk. *The effects of seductive details and segmentation on interest, recall and transfer in a multimedia learning environment*. PhD thesis, Virginia Tech, 2008. → page 57

[41] M. Mavrikis and E. Geraniou. Using qualitative data analysis software to analyse students computer-mediated interactions: the case of migen and transana. *International Journal of Social Research Methodology*, 14(3): 245–252, 2011. → page 18

[42] R. E. Mayer and P. Chandler. When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of educational psychology*, 93(2):390, 2001. → page 7

[43] M. Merkt, S. Weigand, A. Heier, and S. Schwan. Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction*, 21(6):687–704, 2011. → page 57

[44] N. Mirriahi, D. Liaqat, S. Dawson, and D. Gašević. Uncovering student learning profiles with a video annotation tool: reflective learning with and without instructional norms. *Educational Technology Research and Development*, 64(6):1083–1106, 2016. → pages xi, 14, 16

[45] E. S. Mollashahi, M. S. Uddin, and C. Gutwin. Improving revisitation in long documents with two-level artificial-landmark scrollbars. In *Proceedings of the International Conference on Advanced Visual Interfaces-AVI*, volume 18, pages 1–9, 2018. → page 10

[46] F. A. Moretti and H. P. Ginsburg. Video interactions for teaching and learning (vital): A learning environment for courses in early childhood mathematics education. 2009. → pages xi, 15, 17

[47] F. M. Newmann. *Student engagement and achievement in American secondary schools.* ERIC, 1992. → page 1

[48] C. Nguyen and F. Liu. Gaze-based notetaking for learning from lecture videos. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2093–2097. ACM, 2016. → page 18

[49] I. A. Ovsiannikov, M. A. Arbib, and T. H. McNeill. Annotation technology. *International journal of human computer studies*, 50(4):329, 1999. → pages xi, 3, 12, 13, 41

[50] N. Prestopnik and A. R. Foley. Visualizing the past: The design of a temporally enabled map for presentation (tempo). *International Journal of Designs for Learning*, 3(1), 2012. → pages xi, 8, 9

[51] P. Rich, A. Recesso, M. Allexsaht-Snider, and M. Hannafin. The use of video-based evidence to analyze, act on, and adapt preservice teacher practice. In *annual meeting of the American Educational Research Association, Chicago*, 2007. → page 14

[52] A. J. Rockinson-Szapkiw, J. Courduff, K. Carter, and D. Bennett. Electronic versus traditional print textbooks: A comparison study on the influence of university students' learning. *Computers & Education*, 63:259–266, 2013. → pages 3, 12

[53] J.-F. Rouet and B. Coutelet. The acquisition of document search strategies in grade school students. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22(3): 389–406, 2008. → page 57

[54] A. Saxena and R. Stevens. Video traces: creating common space between university and public schools for preparing new teachers. In *Proceedings of the 8th iternational conference on Computer supported collaborative learning*, pages 643–645. International Society of the Learning Sciences, 2007. → pages xi, 17, 18

[55] H. V. Shin, F. Berthouzoz, W. Li, and F. Durand. Visual transcripts: lecture notes from blackboard-style lecture videos. *ACM Transactions on Graphics (TOG)*, 34(6):240, 2015. → pages 9, 13

[56] M. Smith. Dynatext: An electronic publishing system, 1993. → page 12

[57] J. Stoop, P. Kreutzer, and J. Kircz. Reading and learning from screens versus print: a study in changing habits: Part 1–reading long information rich texts. *New Library World*, 114(7/8):284–300, 2013. → pages 3, 12, 59

[58] M. S. UDDIN, C. GUTWIN, and A. GOGUEY. Artificial landmarks augmented media player for video revisitation. → page 10

[59] M. S. Uddin, C. Gutwin, and A. Goguey. Using artificial landmarks to improve revisitation performance and spatial learning in linear control widgets. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pages 48–57. ACM, 2017. → pages xi, 10, 11, 56

[60] E. A. Van Es and M. G. Sherin. Learning to notice: Scaffolding new teachers interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4):571–596, 2002. → page 18

[61] F. Vohle. Relevanz und referenz: Zur didaktischen bedeutung situationsgenauer vi deokommentare im hochschulkontext. *Vorwort zur Idee und zum Thema*, page 165, 2013. → page 13

[62] R. Waller. Functionality in digital annotation: Imitating and supporting real-world annotation. *Ariadne*, (35), 2003. → page 11

[63] G. A. Wright. How does video analysis impact teacher reflection-for-action? 2008. → page 14

[64] S. R. Yussen, A. D. Stright, and B. Payne. Where is it? searching for information in a college textbook. *Contemporary Educational Psychology*, 18(2):240–257, 1993. → page 57

[65] C. Zahn, B. Barquero, and S. Schwan. Learning with hyperlinked videosdesign criteria and efficient strategies for using audiovisual hypermedia. *Learning and Instruction*, 14(3):275–291, 2004. → page 58

[66] D. Zhang, L. Zhou, R. O. Briggs, and J. F. Nunamaker Jr. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & management*, 43(1):15–27, 2006. → page 10

[67] X. Zhang. *Investigation of a quick tagging mechanism to help enhance the video learning experience*. PhD thesis, University of British Columbia, 2017. → pages 46, 58

[68] P. H. Zimmerman, J. E. Bolhuis, A. Willemsen, E. S. Meyer, and L. P. Noldus. The observer xt: A tool for the integration and synchronization of multimodal signals. *Behavior research methods*, 41(3):731–735, 2009. → page 15

Appendix A

Surveys

This chapter contains the pre-experiment survey, the session survey, and the post-experiment survey used in the presented experiment.

A.1 Pre-Experiment Survey

1. I am familiar with ViDex.

○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree

Section I

1. How often do you watch videos?

○ Never ○ Less than a day per week ○ One day per week ○ 2 - 4 days a week ○ 5 - 7 days a week ○ Every day, once a day ○ Every day, more than once a day

Please specify or comment:

_____

2. How often do you watch videos for learning (e.g., how-to instruction, problem solving, facts, etc)?

○ Never ○ Less than a day per week ○ One day per week ○ 2 - 4 days a week ○ 5 - 7 days a week ○ Every day, once a day ○ Every day, more than once a day

Please specify or comment on the frequency and the types of learning:

_____

3. What videos do you watch for learning? (e.g., YouTube videos, MOOC videos, videos from teachers, etc)

_____

(Answer Section II and Section III only if you watch videos for learning sometimes)

Section II

1. How many times do you usually rewatch a video that you have watched for learning before?

○ Never ○ Once ○ 2 - 5 times ○ 6 - 10 times ○ > 10 times

Please specify or comment:

_____

(Answer 2 - 5 only if you did not select “Never” in 1)

2. Why do you rewatch a video that you have watched for learning before? (e.g., for an exam, to learn again, for fun, to kill time, etc)

_____

3. When do you rewatch a video that you have watched for learning before?(e.g., the day before an exam, regularly throughout the term, randomly, etc)

_____

4. How do you rewatch a video that you have watched for learning before?(e.g., rewatch the whole video, rewatch some segments of the video, rewatch randomly etc)

_____

5. How difficult is it for you to find a segment of a video that you have watched for learning before to rewatch?

1	2	3	4	5	6	7
Very easy			Neutral			Very difficult

Please specify or comment on how you find a segment of a video:

_____

Section III

1. How often do you take notes while watching a video for learning?

Never	Once for a video	At least twice for a video	At least once for every topic	At least once for every minute

Please specify or comment:

_____

(Answer 2 - 5 only if you take notes while watching a video for learning)

2. How many times do you usually review your notes on a video that you have watched for learning before?

Never	Once	2 - 5 times	6 - 10 times	> 10 times

Please specify or comment:

_____

3. Why do you review your notes on a video that you have watched for learning before? (e.g., for an exam, to learn again, for fun, to kill time, etc)

_____

4. When do you review your notes on a video that you have watched for learning before? (e.g., the day before an exam, regularly throughout the term, randomly, etc)

_____

5. How do you review your notes on a video that you have watched for learning before? (e.g., review all notes, review notes on some topics, randomly, etc)

_____

Section IV

1. How often do you take notes while reading a book for learning?

Never	Once for a book	At least twice for a book	At least once for every chapter	At least once for every topic	At least once for every page
○	○	○	○	○	○

Please specify or comment:

_____

2. How often do you make highlights while reading a book for learning?

Never	Once for a book	At least twice for a book	At least once for every chapter	At least once for every topic	At least once for every page
○	○	○	○	○	○

Please specify or comment:

_____

(Answer 3 - 6 only if you make notes or highlights while reading a book for learning)

3. How many times do you usually review your annotations(notes or highlights) on a book that you have read for learning before?

Never	Once	2 - 5 times	6 - 10 times	> 10 times
○	○	○	○	○

Please specify or comment:

_____

4. Why do you review your annotations on a book that you have read for learning before? (e.g., for an exam, to learn again, for fun, to kill time, etc)

_____

5. When do you review you annotations on a book that you have read for learning before? (e.g., the day before an exam, regularly throughout the term, randomly, etc)

_____

6. How do you review you annotations on a book that you have read for learning before? (e.g., review all annotations, review annotations on some topics, randomly, etc)

_____

A.2 Session Survey

Section I

1. I knew a lot about the topic before I watched this video.

○ ○ ○ ○ ○

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

2. I was interested in the topic just now.

○ ○ ○ ○ ○

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

3. I was interested in the video just now.

○ ○ ○ ○ ○

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

4. The video is effective in teaching the topic.

○ ○ ○ ○ ○

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

Please specify or comment:

_____

Section II

1. I was focusing more on learning than on making annotations just now.

○ ○ ○ ○ ○

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

Please specify or comment:

_____

2. Learning the video content took me more efforts than making annotations just now.

○ ○ ○ ○ ○

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

Please specify or comment:

_____

3. Making annotations helped my learning just now.

○ ○ ○ ○ ○

Strongly disagree Somewhat disagree Neutral Somewhat agree Strongly agree

Please specify or comment:

_____

4. I learnt very well just now.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
○	○	○	○	○

Section III

1. I am familiar with the content of the video now.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
○	○	○	○	○

Please specify or comment:

_____

2. I would review annotations I made just now for an exam of the course.

Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree
○	○	○	○	○

Please specify or comment:

_____

A.3 Post-Experiment Survey

Section I

1. The annotation manager is easy to use.

○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree

2. I like using the annotation manager.

○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree

3. Using the annotation manager saved me time in finding a video segment compared to not using it.

○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree

Please explain or comment:

_____

4. I would use the annotation manager(if I have one) when I need to rewatch a video that I have watched for learning before.

○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree

Please explain or comment on how you would use the annotation manager:

_____

Section II

1. It was easy to make annotations(highlights, tags, notes) while watching a video for learning just now.

○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree

Please explain or comment:

_____

2. It was easy for me to get used to taking annotations(highlights, tags, notes) while watching a video for learning just now.

○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree

Please explain or comment:

_____

Section III

1. I would review my video annotations(highlights, tags, notes) for my goals such as to pass an exam.

○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree

Please explain or comment:

_____

2. I made annotations(highlights, tags, notes) more for learning than for the experiment tasks just now.

○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree

Please explain or comment:

_____

3. I would make annotations(highlights, tags, notes) for my course video(If I have one on ViDex) in a similar way as I did in the experiment just now.

○ Strongly disagree ○ Somewhat disagree ○ Neutral ○ Somewhat agree ○ Strongly agree

Please explain or comment:

_____

Appendix B

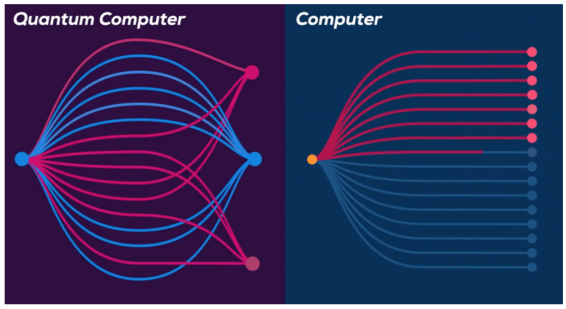
Searching Tasks

B.1 Quantum Computation

【AM】
1. Please find a point when the speaker is talking about:

superposition

2. Please find the point of:



【AM】
3. Please find a point when the speaker is talking about:

entanglement

4. Please find the point when the speaker is saying:

“so as long as it's unobserved, the qubit is in a superposition of probabilities for zero and one, and you can't predict which it will be, but the instant you measure it, it collapses into one of the definite states.”

【AM】
5. Please find the point of:

6. Please find a point when the speaker is saying:

“so a quantum computer sets up some qubits applies quantum gates to entangle them and manipulate probabilities. Then finally measures the outcome collapsing superpositions to an actual sequence of Zeros and ones.”

7. Please find the point when the speaker is talking about:

a normal logic gate VS. a quantum gate

【AM】
8. Please find a point when the speaker is saying:

“Once you can add, you can also multiply and once you can multiply, you can basically do anything.”

9. Please find a point when the speaker is talking about:

the reason why scientists are trying to build quantum computers

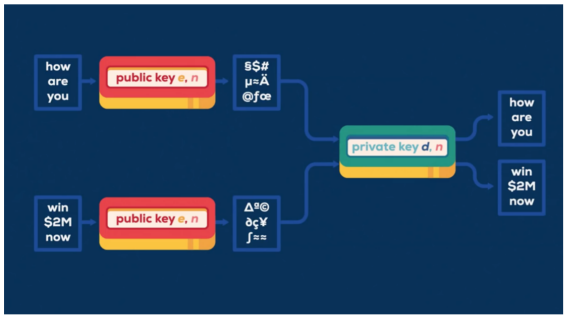
10. Please find a point when the speaker is talking about:

how information is represented in computers

【AM】
11. Please find the point when the speaker is talking about:

quantum computers on database searching

【AM】
12. Please find the point of:



B.2 Fusion Power

1. Please find a point when the speaker is talking about:

nuclear fusion makes the sun shine

【AM】

2. Please find a point when the speaker is talking about:

the safety of fusion reactor

3. Please find the point when the speaker is saying:

“Helium 3, an isotopes of helium, might be a great substitute. Unfortunately it's also incredibly rare on earth but here the moon might have the answer.”

4. Please find the point when the speaker is saying:

“ Fossil fuels are extremely toxic, nuclear waste is, well, nuclear waste and there are not enough batteries to store sunlight for cloudy days yet ”

5. Please find the point when the speaker is saying:

“On Earth, It's not feasible to use this brute force method to create fusion so if we wanted to build or react to that generates energy from fusion we have to get clever”

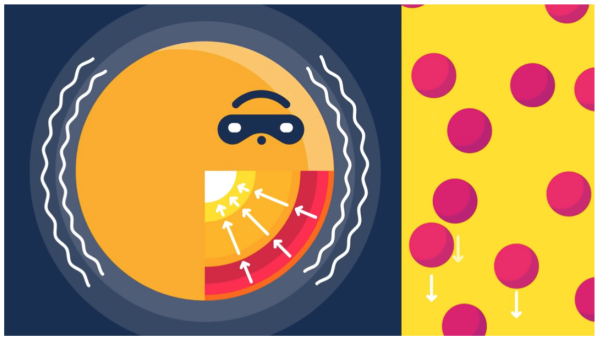
6. Please find the point of:

【AM】
7. Please find a point when the speaker is talking about:

ways we can use to get energy

【AM】
8. Please find the point of:

【AM】
9. Please find the point of:



【AM】
10. Please find a point when the speaker is talking about:

inertial confinement reactor

【AM】

11. Please find a point when the speaker is talking about:

fuel of fusion reactors

12. Please find a point when the speaker is talking about:

magnetic confinement reactor

B.3 The Brain

1. Please find a point when the speaker is talking about:

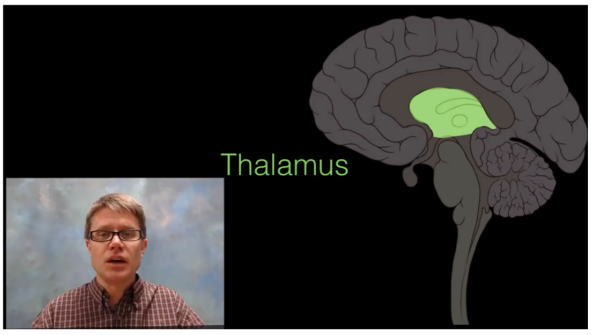
structure of brainstem

【AM】
2. Please find a point when the speaker is talking about:

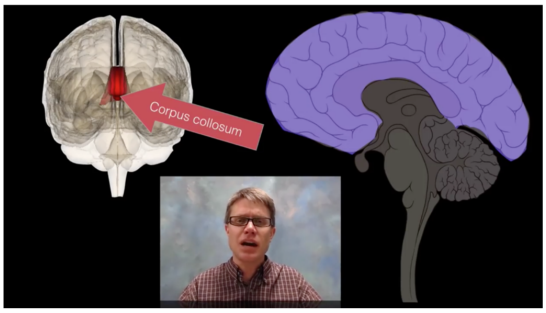
functions of posterior pituitary

3. Please find the point of:

【AM】
4. Please find the point of:



【AM】
5. Please find the point of:



【AM】
6. Please find a point when the speaker is talking about:

the application of functional MRI

7. Please find the point when the speaker is saying:

“what does that accountable for is homeostasis, it's maintaining body temperature, it's maintaining osmolarity, all of that stuff is contained right up in the hypothalamus, also important in circadian rhythms. ”

【AM】
8. Please find a point when the speaker is talking about:

basal ganglia

9. Please find a point when the speaker is talking about:

functions of cerebellum

10. Please find a point when the speaker is talking about:

functions of cerebrum

11. Please find the point when the speaker is saying:

“we do tend to show a lateralization, there are certain things that we put kind of on the left side of our brain, like a mathematical reasoning and logic, and things that we put on the right side, like facial recognition.”

【AM】

12. Please find the point when the speaker is saying:

“The first thing it's gonna do are these more basic needs; it's going to keep your self-breathing, keep circulation going, digestion, swallowing, all of that is gonna be controlled by the brainstem.”

B.4 No AM Condition

【with or without AM】

1. Please find a point when the speaker is talking about:

structure of brainstem

Appendix C

User Study Data

participant	watching_order	video	task_num	item_num	AM	annotated	annotation	t
1	0	0	1	T3	Y	Y	NT	8.73915944
1	0	0	2	02	N	Y	HL	70.8805871
1	0	0	3	T4	Y	Y	HL	13.3422282
1	0	0	4	05	N	Y	NT	24.0160107
1	0	0	5	01	Y	N	N	17.9452969
1	0	0	6	06	N	N	N	46.5310207
1	0	0	7	T5	N	N	N	149.933289
1	0	0	8	04	Y	N	N	23.9492995
1	0	0	9	T2	N	N	N	37.5917278
1	0	0	10	T1	N	Y	TG	9.50633756
1	0	0	11	T6	Y	Y	NT	8.30553703
1	0	0	12	03	Y	N	N	20.9472982
1	1	1	1	T1	N	Y	NT+HL	20.1801201
1	1	1	2	T3	Y	N	N	33.5557038
1	1	1	3	02	N	N	N	18.245497
1	1	1	4	01	Y	Y	HL	8.83922615
1	1	1	5	03	Y	N	N	10.7404937
1	1	1	6	T5	Y	Y	NT	5.17011341
1	1	1	7	05	N	Y	HL	15.7438292
1	1	1	8	T6	Y	Y	NT	6.80453636
1	1	1	9	T2	N	Y	HL	7.80520347
1	1	1	10	T4	N	Y	HL	8.57238159
1	1	1	11	06	N	N	N	15.6771181
1	1	1	12	04	Y	N	N	35.323549
1	2	2	1	T2	N	N	N	8.00533689
1	2	2	2	T6	Y	Y	HL	6.43762508
1	2	2	3	06	Y	Y	HL	5.30353569
1	2	2	4	04	N	Y	NT	6.8712475
1	2	2	5	05	N	Y	HL	14.9432955
1	2	2	6	02	N	N	N	17.6784523
1	2	2	7	T1	N	Y	NT	4.43629086

Appendix D

R Code

```
data<-read.csv("formal_exps_aggregated.csv")
s0 <- subset(data, watching_order==0)
s1 <- subset(data, watching_order==1)
s2 <- subset(data, watching_order==2)
library(lme4)
library(lmerTest)

### session 0
s0_item = lmer(t~AM + annotated + AM:annotated + task_num +
                 AM:task_num + (1|participant) + (1|item_num),
               REML=TRUE, data=s0 )
summary(s0_item, ddf="Kenward-Roger")
anova(s0_item, ddf="Kenward-Roger")

### session 1
s1_item = lmer(t~AM + annotated + AM:annotated + task_num +
                 AM:task_num + (1|participant) + (1|video/item_num),
               REML=TRUE, data=s1)
summary(s1_item, ddf="Kenward-Roger")
anova(s1_item, ddf="Kenward-Roger")

### session 2
s2_item = lmer(t~AM + annotated + AM:annotated + task_num +
                 AM:task_num + (1|participant) + (1|video/item_num),
               REML=TRUE, data=s2)
summary(s2_item, ddf="Kenward-Roger")
anova(s2_item, ddf="Kenward-Roger")
```

Appendix E

Power Analysis

The power analysis was performed using the software G*Power, based on a pilot study with 4 subjects.

The pilot study was a simplified version of the formal experiment, in which all 4 subjects watched the 3 videos in the same order across 3 sessions (Quantum Computation, The Brain, Fusion Power), and finished the same set of searching tasks with identical AM conditions in session 1.

Within-subjects analysis was performed to obtain the partial η which was needed to determine the effect size f. Means of subjects' task performances in session 1 are shown in Table E.1. The AM condition was treated as the within-subjects factor, and the within-subjects partial η was 0.67. The effect size f was then calculated by G*Power.

Parameters used in the power analysis are shown in Table E.2, and a total sample size of 4 was recommended by the power analysis.

Subject ♯	AM	Non AM	Annotated	Non Annotated
1	33.60	47.73	32.63	64.78
2	13.42	20.92	15.07	27.70
3	15.70	16.56	15.08	27.75
4	19.54	41.20	29.10	36.76

Table E.1: Means of subjects' task performances in session 1 in the pilot study (measured in seconds).

Parameter	Value
Test family	F tests
Statistical test	ANOVA: Repeated measures, within factors
Effect size f	1.428
α	0.05
Power	0.95
Number of groups	2
Corr among rep measures	0.5
Nonsphericity correction	1

Table E.2: Parameters for power analysis on G*Power.