Investigation of a Quick Tagging Mechanism to Help Enhance the Video Learning Experience

by

Xueqin Zhang

B.E., The University of Electronic Science and Technology of China, 2014

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)

December 2017

Abstract

Video continues to be used extensively as an instructional aid within modern educational contexts, such as in blended (flipped) courses, self-learning with MOOCs (Massive Open Online Courses), informal learning through online tutorials, and so on. One challenge is providing mechanisms for students to efficiently bookmark video content and quickly recall and review their video collection. We have run a background study to understand how students annotate video content, focusing especially on what words they would use most to bookmark video content. From this study, we proposed to leverage a quick tagging mechanism in an educational video interface comprised of a video filmstrip and transcript, both presented adjacent to a video player. The 'quick' tagging is defined as an easy and fast way to mark course video parts with predefined semantic tags. We use the metaphor of marking and highlighting textbook to achieve our quick tagging interaction. This mechanism was evaluated in a controlled study with handwritten notes. We found that participants using our quick tagging interface spent around 10% longer watching and learning from video on average than when taking notes on paper. Our participants also reported that tagging is a useful addition to instructional videos that helps them recall video content and finish learning tasks.

Lay Summary

Video is widely used as an instructional aid within educational contexts such as blended (flipped) courses, self-learning with MOOCs (Massive Open Online Courses), informal learning through online tutorials, and so on. One challenge we face is to support students with efficient video content bookmarking and to improve their recall and review of their video collection. We have run a background study to understand how students use texts to summarize or record video content, especially focusing on what words they use most to bookmark video content. From this study, we proposed to integrate functions in an educational video interfaces, such as adding tags, deleting tags, and using tags to jump around while watching the video. Here, tags are a set of key words. The interface is comprised of a video filmstrip and transcript, both presented adjacent to a video player. A video filmstrip is a set of thumbnails from the video arranged side by side, each representing a portion of the video. Our interface including the aforementioned tagging functions was compared with taking notes on paper in a lab study. Tagging refers to the actions users take to manipulate tags, such as adding tags, deleting tags and editing tags. We found out that participants using our interface spent around 10% longer watching and learning from video on average than when taking notes on paper. Our participants also reported that such tagging functions are useful additions to instructional videos to help them quickly recall video content and finish learning tasks.

Preface

All of the research work presented in this thesis was conducted in the Human Communication Technologies Laboratory (HCT) at the University of British Columbia, Point Grey campus. All user studies and associated methods were approved by the University of British Columbia Behavioural Research Ethics Board [certificates #: H13-01589].

All of the implementation and experiments henceforth were conducted by myself. Concepts and design decisions were discussed among myself, Matthew Fong, Gregor Miller and Sidney Fels.

A version of Chapter 4 and Chapter 5 was submitted as Zhang, X.Q., Miller, G, Fong, M, Roll, I, Fels, S (2018) at CHI 2018.

Table of Contents

Appendix

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank my loving family. Without them, I may have given up. Of course, I also want to thank my supervisor, Dr. Sidney Fels, for overseeing this work, as well as Dr. Gregor Miller and Matthew Fong, who worked with me on much of this work. Finally, I would like to thank my best friends, Qingqing and Marco who have also always accompanied me when I needed support most.

Chapter 1

Introduction

Video has a long history, constituting a vivid and entertaining way of stimulating a learner's interest, conveying desired knowledge, and being used as an educational tool, such as through documentaries and visual tutorials. In the past decade, free online video hosting services, such as YouTube, have led to a significant rise of varied educational videos across many topics, contributed to by individuals and well-known universities alike. Videos have become central to the student learning experience in the current generation of Massive Open Online Courses (MOOC) from providers such as Coursera, edX and Udacity. In 2014, at least 20,000,000 learners registered to at least one MOOC[1]. These online education frameworks offer complete courses delivered over the web and video has become a core aspect for self-learning. For learners who seek information on specific topics, sources such as Khan Academy offer more specifically targeted videos.

An empirical study of MOOC videos by Guo et al. [12] indicated that video production style often affects student engagement. They summarized four typical styles, including classroom lectures, the "talking head" shot of an instructor, digital tablet drawing format, and PowerPoint slide presentations. Some interesting findings from their work revealed that shorter videos are much more engaging and that students engage in different ways with lecture and tutorial videos. In this paper, video engagement was measured by how long students spend watching each video and whether they attempted to answer post-video assessment problems. Lecture videos usually present conceptual (declarative) knowledge, whereas tutorials present how-to (procedural) knowledge (e.g., problem solving walkthrough). The findings of Guo et al. suggest that students expect a lecture to be a continuous stream of information, so instructors should provide a good experience for first time watchers. For tutorials, re-watching and skimming should also be supported. While issues can be solved by investing heavily in pre-production lesson planning, providing an efficient mechanism for learners to facilitate their video learning process to keep them engaged can further enhance the

[1] https://www.edsurge.com/news/2014-12-26-moocs-in-2014-breaking-down-the-numbers

solution.

Five typical video watching scenarios under learning contexts are summarized by Kim et al. [15] including the following: re-watching to better understand a concept, textual searching for a specific phrase mentioned by the instructor, visually searching for a code example scene, returning to a specific slide, and skimming a trivial lecture. To support learners, the researchers fed patterns of interaction data back into the video navigation interface. Although the collective interaction data might indicate points of learners' interest, confusion, or boredom in videos, the data-driven techniques might also ignore other potentially important points. Specifically, the interaction data in their work is limited to clickstream logging, and other types of interaction data like active bookmarking and content streams (such as voice) should also be explored to obtain comprehensive results.

There are other works trying to approach the issue of supporting learners' navigation and recalling video content. For facilitating navigation of blackboard-style lecture videos, Monserrat et al. [21] developed and evaluated *NoteVideo* and its improved version, *NoteVideo+*, systems that support in-scene navigation and directly jumping to a video frame instead of navigating linearly through time. More specifically, the systems automatically identify the conceptual 'object' of a blackboard-based video and then create a summarized image of the video and using it as an in-scene navigation interface that allows users to directly jump to the video frame where that object first appeared. However this solution is limited to a specific video format, and may not work well for less visual information video formats, like "talking head" videos. Further, video digests was proposed by Pavel et al. [25] to help viewers browse and skim long informational talks, lectures, and distance-learning videos online. The key insight of their approach was that much of the information in lecture videos is conveyed through speech. Therefore, they presented a set of tools to help video authors create digests by segmenting videos into a chapter/section structure and providing short text summaries and thumbnails for each section using a time-aligned transcript of the speech. However, creating a video digest is a time-consuming process for authors. They also provided algorithmic tools for automatically segmenting the video and a crowdsourcing pipeline for summarizing the resulting segments so that authors can further refine these auto-generated digests, if necessary. One limitation of this work is that crowdworkers may not have the necessary background to write summaries for highly-technical content.

Figure 1.1 indicates users' mental models in a video learning scenario. Learning goals or tasks motivate users to manipulate video content (such

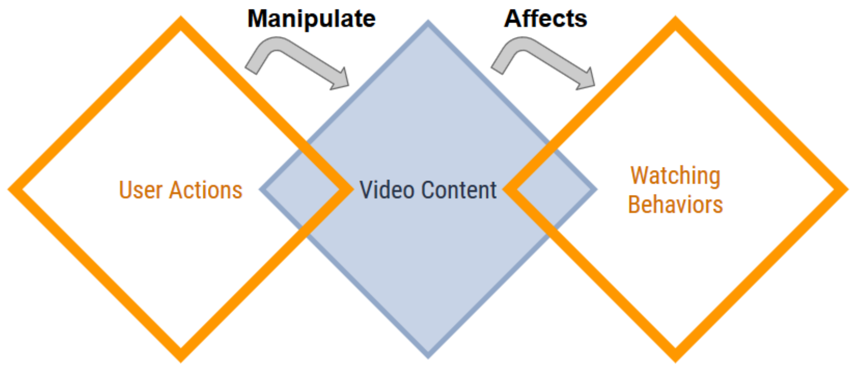

Figure 1.1: Mental model in a video learning scenario

as skimming, re-watching, etc.). Meanwhile, video content affects users' watching behaviors, such as engagement. Based on this mental model, we propose a quick tagging mechanism leveraged on a video interface to help students manipulate video content such as searching, chaptering and recalling as they learn, aimed for enhancing their video learning experience. The "quick" here is an absolute concept instead of a relative one. We are not focused on comparing our quick tagging mechanism with other tagging systems. We believe that every tagging system is well designed to satisfy its own system requirements. As for video learning applications, the "quick" tagging is designed to facilitate the process of manipulating video content and recording useful learning information. It is a way to help students focus on learning while efficiently recording learning content. The quickness is a deterministic element which is amplified from three perspectives. We provide pre-defined tags for users to remove the hinder of being unable to think of any tag. The process of tagging video content adopts the familiar metaphor of marking textbooks. The tagging interaction is well adapted from user habits of highlighting pdf files. Thus, we measure the "quick" from two metrics: efficiency and usefulness. Efficiency means that users can easily use our quick tagging interface with one-click efforts. Usefulness means that our quick tagging mechanism is helpful to qualitatively improve their learning experience and performance.

Searching, chaptering and recalling video content is especially difficult with educational videos, as such videos are generally visually similar. Therefore, it is important to find a way to help learners recognize and remember major topics in video clips. In the learning process, it is common to come across interesting or confusing content, and it is thus helpful for learners to record their current mental state for later review.

Tagging content using keywords is widely used on the web, in social media and in applications such as photograph organization, taking shape in various forms such as tag clouds and clickable search terms. Tags are used for search and recall of content, and as a method for giving feedback on content (often used in social media). Within educational contexts, tags can be used by students to help them search for their content at a later date during their studies, aid them in recall (e.g. why they viewed this video) and provide feedback to the instructor.

Tagging systems for expressing preference typically take the form of a unary mode ("like"/"dislike"), multi-valued (star ratings), or text-based. Each system must make a tradeoff between amount of cognitive load required and the ability to express a full range of reactions. More specifically, for text-based tagging systems, there is a tradeoff between wide tag vocabulary and restricted tag vocabulary. In other words, there is a balance between flexibility and usability that must be found. Although a restricted tag vocabulary sacrifices flexibility to some degree, it can be more easily summarized by automated methods, filtering, and searching and it has a lower cognitive load. A restricted version can be best if a strong method for defining tags is used, such as maximizing tag concept range and limiting tag number.

Brame [4] demonstrated the Cognitive Theory of Multimedia Learning which gives rise to several recommendations about educational videos. One of the recommendations is to highlight important information (called Signaling) which can reduce extraneous load and enhance germane load. In Cognitive Load Theory, extraneous load refers to cognitive effort that does not help the learner toward the desired learning outcome. Alternatively, germane load refers to the level of cognitive activity necessary to reach the desired learning outcome - e.g., to make necessary comparisons, complete analysis, and elucidate the steps necessary to master a lesson. Establishing predefined tags that are representative of active learners' reactions can be useful for video learning. Such predefined tags can be brief out-of-video text which can help explain purpose and context for the respective video. In addition, representative tags can maximize tag concept range, motivate students to tag, and alleviate the tagging hinderance. Sen et al. claim that 68% of non-taggers in their study [33] did not tag because they simply could not think of any tags. If we can provide some useful tags in the tagging interface, this tagging hinderance may be alleviated. Two different approaches to creating a useful collection of tags were suggested by Velsen et al. in their incorporating user motivations to design for video tagging study [37]. One was to motivate all users to tag resources from a system launch on. The other one was to let professionals tag resources until the audience at large is familiar with tagging and more interested in tagging resources. For the latter approach, it was costly to exhaust all the professionals of every course subject. In addition, we cannot ensure the word choices of professionals lead to interesting resources or are easy to understand for the 'average' user.

Meanwhile, an optimal number of tags should be decided in order to build a well restricted tag vocabulary. If the number of default tags is too large, it will require a larger cognitive load for taggers to determine useful tags. If the number of tags is too small, it will not cover the full range of reactions, which thus cannot be fully representative and can cause users not to use the predefined tags.

Reactions to video content can also be highly sensitive to studying contexts. In social and voluntary posting contexts such as YouTube, learners are free to react however they like to video material. But in a context where students are encouraged to tag (ie. under instructor supervision), students may hide their real reactions from their instructors. In light of this, in our work, we focus on collecting one-word tags from two datasets, namely YouTube and the Collaborative Learning Annotation System (CLAS) [29], to determine how best to help learners recall video content. CLAS is a media player used to record, share, and comment on videos. Our main goal is to establish useful predefined tags representative of the reactions of active learners, and to decide which default tag number is optimal. Further details are described in our background study in Chapter 4. Besides predefined tags, the "quick tagging" of our tagging mechanism is also amplified from video part selection and applying tags. More details about interface design strategies for quick tagging educational videos are described in Chapter 3. To explore the design space of the quick tagging interface, we build a web prototype and evaluate it in a controlled study, which is described in Chapter 5.

1.1 Research Question

The purpose of this work is to apply the quick tagging mechanism to a video learning context, meanwhile keeping it simple enough so that learners can efficiently complete common learning tasks. To direct this work, we propose the questions:

1) Do users feel it is efficient and useful enough to perform quick tagging on video content while finishing their learning tasks? 2) In perspective of usefulness, does quick tagging mechanism help students recall video content?

1.2 Contributions

We focus our efforts on answering our aforementioned research question in a learning context. We developed a new video viewing and navigation interface designed to integrate the quick tagging mechanism for efficiently bookmarking video content. We ran one background study, two pilot studies, and one controlled lab study to collect predefined tags, test the design, and verify its efficiency and usefulness. Thus, the main contribution of this work is to explore and investigate how a quick tagging mechanism enhances the user learning experience from the perspective of efficiency and usefulness in recalling video content.

In this work, there are two levels of contribution. The first is a background study comprised of content analysis and an online survey. This study gave us insights on what tags are commonly used by learners to describe video content and how the polarity of reaction words can be biased by learning contexts. The second contribution is investigating user preference and watching patterns on our quick tagging interface, comparing user performance in the quick tagging interface with hand-written notes to finish a quiz related to the video content. Through the user study, we demonstrate the effectiveness of the quick tagging interface where users bookmark video content as well as its usefulness in improving video content recall. We also found implications for further improving the quick tagging mechanism in aspects of the predefined tags, the quick tagging interaction, and the interface visualization.

1.3 Publications

At the Graphics Interface Conference (2016), we published one paper. The paper [10] designed an interface which uses textbook-style highlighting on a video filmstrip and transcript, both presented adjacent to a video player. The qualitative results indicated that the familiar interaction of highlighting text was preferred, with the filmstrip used for intervals with more visual stimuli. My main contribution was to help conduct a preliminary investigation and qualitative user study together with G. Miller and M. Fong. I also contributed to improving the highlighting interface by taking part in the design meeting.

Together we submitted a paper for CHI 2018 which covered the work of Chapter 5 and part of Chapter 4.

Chapter 2

Related Work

In this chapter, we first introduce the definition of tags and tagging in relation to their functions, advantages, and applications. Our mechanism of using predefined tags for bookmarking educational video content efficiently relies on tag taxonomy methods, video annotation, and video interface design for education. The overview of taxonomy looks at methods for synonym-based word frequency analysis and various verb classifications. We take methodology cues from the presented works to integrate into our own methods for tag taxonomy, which considers the learning context factor. There are currently several video annotation tools being developed, allowing users to mark up video in personalized or collaborative ways, and forming a spectrum from manual to fully-automatic that involves different methods of annotations, such as text, ink, and link. Lastly, the way we choose the video interface features and elements is based on the respective context of educational video content and learning.

2.1 Tags/Tagging

There is a rich thread of research in studying the functions and advantages of tags and tagging. Scott et al. [11] identifies seven functions that tags perform for bookmarks. The authors mention that most tags identify the topics of bookmarked items, and also indicated that adjectives such as “scary”, “funny” and “stupid” were inspirational tag bookmarks according to the tagger’s opinion of content. Collaborative tagging [19] is a practice whereby users assigned uncontrolled keywords to information resources. Social tagging [36] is defined as the collaborative activity of marking shared online content with keywords, or tags, as a way to organize content for future navigation, filtering or searching. Storey et al. [34] investigated whether combining waypointing and social tagging is a useful metaphor to support navigation in the software space of source code and related artifacts. In software space, the authors expected waypoints to be locations of software model elements, or locations that correspond to a file name and line number for any type of file, or for any version of a file. Here, waypoints are indexed

Figure 2.1: The user interface of an Interactive Shared Education Environment, by Mu [23]

through a set of tags supplied by programmers. Further, in a tagging survey [33], it was found that most users who use tags think that tagging features help them express opinions and organize movies.

Tags have been widely used in learning tools, social document annotation systems and social media. In the data-driven interaction design for educational videos [15], word clouds were used to help learners recognize and remember major topics in a video clip. In a social document annotation system study [40], Zyto et al. argued that some comments used lots of screen real estate to convey small bits of information, sometimes obscuring more substantive information. In their results of examining the comments of 5 words or less, they found that 2.7% of the total comments could have been replaced by 8 tags without loss of meaning. Tagged and Hashtagged are two of the verbs in a unified naming schema in the Connected Learning Analytics (CLA) toolkit [16] which enables data to be extracted from social

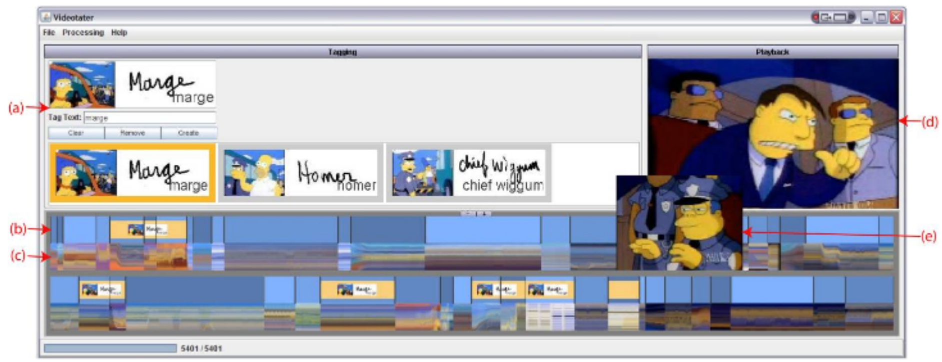

Figure 2.2: The user interface of Videotater, by Diakopoulos et al.

media to facilitate a set of learning activities. Niemann [24] presented a new way to automatically assign tags and classifications to learning objects offered by educational web portals, to address the issue of where data sets often suffer from sparsity when seen in the educational domain. Han et al. [13] aimed to understand the formation of a social network based on two dimensions - Tags and Likes. In the network, the authors used tags to find people with the same interest, and Likes to find connections among those people.

Pina et al. [26] used synonym-based word frequency analysis to develop a taxonomy of public health quality improvement concepts. They analyzed public health-related documents for word frequency, identified the most frequently recurring word-meaning clusters, and created high-level categories based on the ranked synonym cluster. In this paper, our goal is to find most frequently-used words, with three high-level categories in mind. So, to adapt our goal, we have used the last two steps of the method presented by Pina et al. Korhonen et al. [17] presented a substantial extension to Levin's taxonomy which incorporates 57 novel classes for verbs. The Lexical-semantic verb classifications have proved useful in supporting various natural language processing (NLP) tasks. Sabine [31] investigated whether human associations to verbs, i.e., the words that are called to mind by stimulus verbs, as collected in a web experiment, can help us to identify salient features for semantic verb classes. These methods only target verbs and classify words based on human associations. In our work, we target a wider range of words, such as verbs, nouns, adjectives, etc.

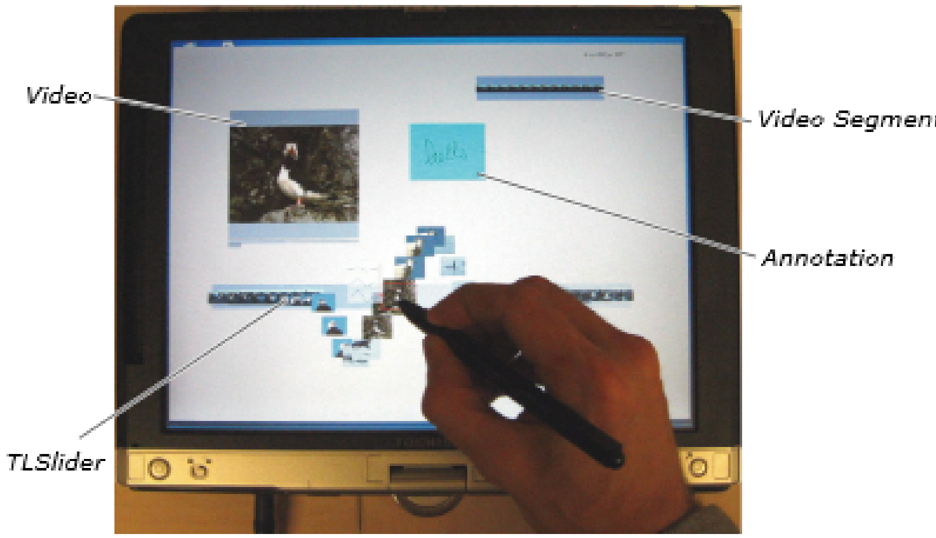

Figure 2.3: The user interface of *LEAN*, by Ramos et al.

2.2 Video Annotation

It is difficult to translate the characteristics of annotations, such as highlighting, context-based notes, and organization from the traditional paper-based medium to a time-based digital format. A video annotation tool called Interactive Shared Education Environment (ISEE) [23], which automatically generates hyperlinked timestamps to associate notes with video content, was developed to explore issues in video annotation. Figure 2.1 illustrates the overview of ISEE's graphic user interface. The system is composed of four components to support different video-based collaborative learning scenarios. Here we focus on the Interactive Chat/Annotation Room (ICR) which is located on the bottom right. The ICR records a clickable link (Smartlink) for each annotation and automatically associates it with a video segment. Clicking one of these Smartlinks navigates the video forward or backward to that specific timestamp and starts playing the video from there.

To examine users' notes-taking behaviors in both individual and collaborative distance learning environments, the usability of ISEE was tested by an empirical comparison study. The results showed that the tool facilitated users' in situ video annotation by allowing users to directly connect their notes with video context. The results also indicated that annotation de-

Figure 2.4: The user interface of the Family Video Archive, by Abowd et al.

pends on the content of information and cues from the lecturer as to what information is important. These findings have motivated us to help users connect their tags with video context so that they can quickly recall video content by referring to key words such as "difficult", "confusing" and "interesting", as well as topic tags. Indeed, the principles of note-talking style change over time, and we believe that our quick tagging mechanism can be considered complementary to traditional note-taking styles.

The evaluation method from an early paper from Bargeron et al. [2] further inspired us. Specifically, their first study of annotation creation on streaming video explored the use of Microsoft Research Annotation System (MRAS) for taking personal notes compared with hand-written notes. MRAS is a web-based client/server application which supports video annotation creation and threaded discussions. This work introduced video annotation in an early stage. Similarly, our work introduces tagging as an annotation tool for educational video in an early stage.

A think aloud test on the usefulness of VideoJot [28] indicated that

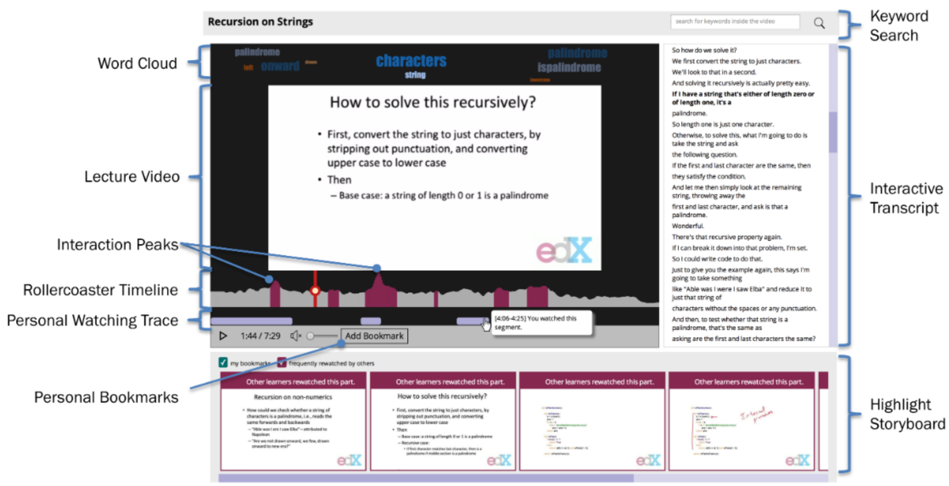

Figure 2.5: The user interface overview, by Kim et al.

both a text annotation tool, as well as a spatial annotation tool are necessary to fulfill the complete information needs of video annotation. They also compared still annotations with moving ones to explore the issue of annotating complex scenes in the video. In addition, LikeLines in VideoJot, a one-dimensional heatmap, was recommended to bookmark interesting parts of long videos. Although their evaluation focused on live and recorded video streams of people playing video games, this work motivated us to think about how to adopt these available video annotation features into our quick tagging interface, especially when educational videos include complex scenes with dense information, such as formula explanations.

Videotater [7], an experimental tool for a Tablet PC, supports the efficient and intuitive navigation, selection, segmentation and tagging of a video. To specifically facilitate the tasks of rapid manual segmentation and tagging, the visual scent of the underlying pixel colors and their evolution on the timeline immediately signals to the user where appropriate segment boundaries should be placed. In addition, rapid review and refinement of manually or automatically generated segments is also supported. A distribution of modalities in the interface by using multiple timeline representations, pressure sensing, and a tag painting/erasing pen metaphor was also explored. As shown in Figure 2.2, the timeline displays two different views of the underlying video, namely timeline segments (b) and timeline strip

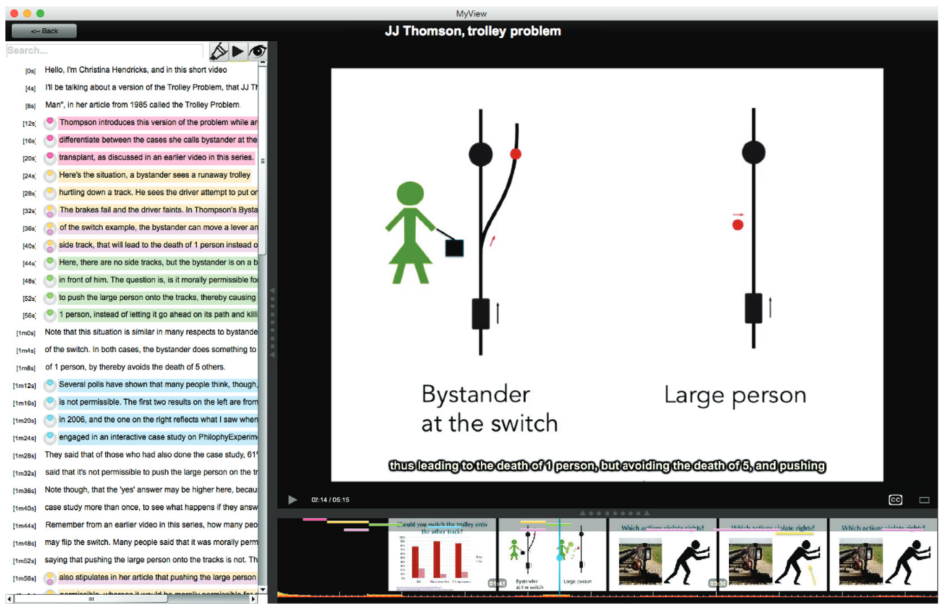

Figure 2.6: The highlighting user interface, by Fong et al.

images (c). The tagging paint is selected from the tagging view (a) and drawn over the timeline wherever it should be applied. When drawn on a segment (b), the tag is applied to the entire segment, whereas drawing on the strip image (c) applies the tag only to the frames touched. This paper is a good example of showing how actions such as video navigation, selection, segmentation and tagging are coupled to support a better user experience of video annotation.

Another paper from Ramos et al. [27] shares the same design principle. A variety of fluid interaction and visualization techniques for navigating, segmenting, linking, and annotating digital videos using a pressure-sensitive pen-based interface are demonstrated within a concept prototype called *LEAN*. An overview of the system can be seen in Figure 2.3, which is composed of Video, Video Segment, Annotation, and TLSlider.

On the opposite end of the spectrum is the fully-automatic approach. The thesis by Morris [22] explores the discovery and effective use of automatic-semantic tags for navigating through video key frames in the unstructured video presentation domain. Machine learning is used to automatically generate classifiers for selected visual concepts. game. Meanwhile, Morris created a new video browser called VastMM-Tag to provide users with data gener-

Figure 2.7: Video Digests, by Pavel et al.

ated by the semantic tagger.

As automatic solutions do not always suit the criteria, we investigate a semi-automatic video annotation tool. The informal nature of home movies makes it difficult to use fully-automated techniques for scene detection and annotation. The Family Video Archive [1] explores the symbiosis between automated and manual techniques for annotation, as well as the use of a zooming interaction paradigm for browsing and filtering large collections of video scenes. A screenshot of the annotation interface is shown in Figure 2.4. The interface mainly supports the annotation of the currently viewed scene. The upper-middle panel of Figure 2.4 shows all metadata associated with the current scene. There are three kinds of annotations possible: date, freeform text, and a metadata tag.

Instead of a personalized solution, an aggregated method also makes sense in certain scenarios. An approach using crowd-sourcing information streaming to semantically annotate live broadcast sports games, and selecting video highlights from the game with these annotations was proposed by Tang et al. [35]. The selected highlights are specific for fans of each team, and these clips reflect the emotions of a fan during a

The underlying concept of the Collaborative Lecture Annotation System (CLAS) [29] is that CLAS moves away from user-defined annotation tools (such as note taking tools) and toward a system-defined annotation strategy; e.g. the annotation has the same meaning for all users. This is a medium through which students and instructors can engage in meaning-making around course content, record important moments, and once finished, navigate these important moments. CLAS relies on semantically-constrained annotation, post-annotation data aggregation, and transparent

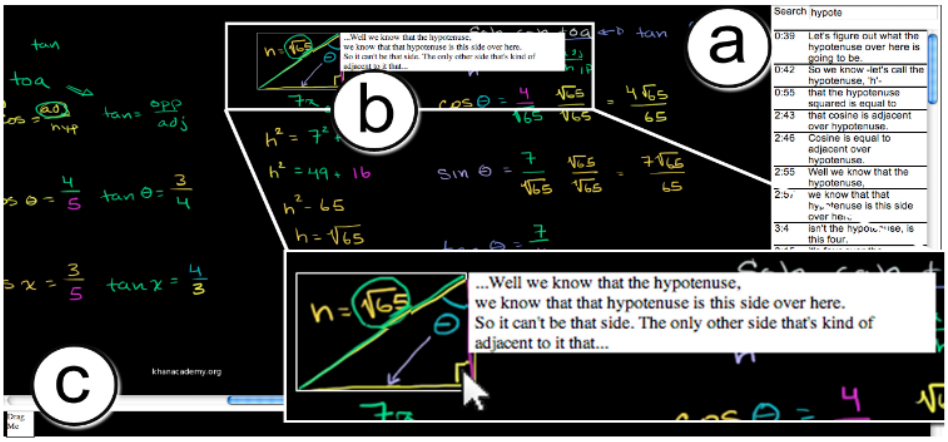

Figure 2.8: *NoteVideo+*, by Monserrat et al.

display of this aggregated data. Our system is not designed as an alternative to CLAS, but instead can be considered as complementary to CLAS's semantically-constrained annotation. Indeed, our system also helps students record important moments, manage their video content, and navigate these important moments while also providing varied approaches in order to facilitate user tagging behaviors while they watch and learn from videos. Generated tags can be further aggregated in a collaborative annotation system.

The Rich Interactive Multimedia Exercise System (RIMES) [14] is a system for authoring, recording and reviewing interactive multimedia exercises embedded in video lectures. In their RIMES creation experience study, researchers coded all RIMES exercises by subject area, by knowledge type, and by the level of cognitive process, using the revised version of Bloom's Taxonomy [18]. The original Bloom's Taxonomy [5], as well as the revised version [18], are common frameworks used to evaluate the level of learning assessed in exams and tests [39]. In our work, we used the guitar video content and a quiz to simulate the learning process from the remember to the analyze level based on the original Bloom's Taxonomy [5].

2.3 Video Interfaces for Education

Free online education video platforms such as Khan Academy, Coursera, edX, Udacity, MIT OpenCourseWare, and YouTube are watched by millions of people. With this unprecedented scale of educational video consumption, there are many challenges and opportunities for us to explore in the design space of educational video interfaces.

Data-driven interaction techniques for educational video navigation are explored by Kim et al. [15]. This study summarizes a few typical video watching scenarios in learning contexts, including: rewatch, textual search, visual search, return, and skim. A set of techniques that augment existing video interface widgets are also included. As shown in Figure 2.5, three sets of novel interaction techniques are presented in the above study. The dynamic timelines are comprised of the following: Rollercoaster Timeline, Interaction Peaks, and tracing of personal watching. In-video searching is enhanced by keyword searching and interactive transcript. Highlights are made with Word Cloud, Personal Bookmarks and the Highlight Storyboard.

Traditional playback controls (play, pause, seek) do not suit well for supporting recall, history, or interval bookmarking. An interface which uses textbook-style highlighting on a video filmstrip and transcript, both presented adjacent to a video player, was designed by Fong et al. [10] for students to manage their video collection and quickly review or search for content. The main video player view is shown in Figure 2.6. According to their experiment results, including transcripts in the videos offers more utility for users and will allow them to highlight, search, and review videos more easily. This work inspired us in the design of our quick tagging interface.

It is difficult to browse and skim the content of long informational lectures using current timeline-based video players. In response, video digests [25] are a new format for informational videos that allow browsing and skimming by segmenting videos into a chapter/section structure and providing short text summaries and thumbnails for each section. A set of tools have been presented to help authors create digests using transcript-based interactions. As shown in Figure 2.7, there are two panes in the interface. One is an aligned transcript pane on the right for navigating, segmenting and summarizing the talk. The other is a WYSIWYG editor pane on the left for adding chapter titles, summaries, and keyframes for each section. Additionally, a progress overview scrollbar in the transcript pane allows authors to view their segmentation progress and return to areas for refinement.

The typical goals of students, like quickly finding a particular concept in a blackboard-style lecture video, are not adequately supported in current video navigation tools. An improvement based on *NoteVideo*, *NoteVideo+* [21] is a system for identifying the conceptual 'object' of a blackboard-based video and then creating a summarized image of the video to be used as an in-scene navigation interface that allows users to directly jump to the video frame where that object first appeared, rather than navigating linearly through time. In Figure 2.8,*NoteVideo+* adds a search box for the transcript (a), a hovering transcript text over the visual elements (b), and a scrubber (c) to address two limitations of the previous interface.

2.4 Summary and Influence on Design

In this chapter, we have covered tags/tagging, video annotation, and educational video interfaces. In the tags/tagging section, we reviewed the following tags/tagging functions and advantages: identifying topics of items or expressing taggers' opinions[[11], [33], [15]], organizing content for future navigation, and filtering or searching[[36], [34], [33], [24]]. We see that tags can be extracted from a large quantity of comments [40] or social media data[[16], [13]] to support learning activities or help understand social network relationships. We also reviewed a word frequency analysis taxonomy method [26] and semantic verb classifications in supporting various natural language processing (NLP) tasks[[17], [31]]. In the video annotation section, we have gained insights on how video navigation, segmentation, and annotation are associated with one another. We also learned that user notes-taking behaviors are different between personal and collaborative learning environments. Finally, we investigated different types of educational video interfaces, which had direct implications on the choices of video interface features and elements.

Related works to this study show the promise of integrating tagging features into educational video interfaces to support learning activities. For varied formats of course video content (such as talking heads, long informative videos, blackboard-style lectures and so on), tags can be useful for identify course topics, expressing learners' opinions, and helping recall video content. Additionally, it is important to consider tagging valence (such as having all positive tags or having a range of tags) for different learning environments. Similar to early works of video annotation[[23], [2]], we have compared our quick tagging mechanism with traditional notes-taking methods as a starting point to investigate the usefulness and efficiency of our quick tagging mechanism in educational video interfaces.

Chapter 3

General Interface Design for Quickly Tagging Educational Videos

Based on our pilot study with eight student participants at the University of British Columbia, we developed an interface that supplies users with four basic functions: watching video, multiple video navigation ways, tagging video parts with two orders, and deleting tagging applications in "un-tagging" mode. The important components in our design are: the quick tagging mechanism (including pre-defined tags, video part selection, and applying tags or vice versa) and a video player adjacent to a video filmstrip and transcript. In this chapter, we describe the complete video interface that integrates our components with features from typical online educational platforms. We also rely on some of the design decisions found by Fong et al. [10].

3.1 Groundwork

The preliminary investigation by Fong et al. [10] informs us that being able to select intervals of video for manipulation is useful for users to manage their video collections. As found in the study, users wanted more control over the ability to emphasize or de-emphasize certain video parts without having to watching the video over and over again. It was stated that being able to see which parts of the video needed more attention would be beneficial in the user's reviewing process.

As shown in Figure 2.6, we decided to use Fong et al.'s three major elements: the player, the filmstrip, and the subtitle viewer. For our study we shared their design idea that allows users to manipulate emphasis on certain video parts by tagging and to use these tags to review video content. We also shared their metaphor of textbook-style tagging for educational video material. More specifically, we supported user tagging on both the filmstrip and transcript. We note that our work is not an alternative to the highlighting system of Fong et al., as tagging can be integrated into their highlighting system. From one perspective, tags can provide more information than highlighting colors; in general, highlighting colors can be integrated into one "important" tag. However, highlighting does afford greater visualization stimuli than tags. From another perspective, our quick tagging shares the same manipulation interaction with highlighting. In the study of Fong et al., their results showed that highlighting on a transcript was preferable to highlighting on a filmstrip. Thus, we wanted to explore whether this finding could be applied to our quick tagging method.

3.2 Use Case

Alice is a first year college student in Electrical and Computer Engineering. This term, she starts learning an entry-level electronic circuit course. One day, she is asked by her instructor to preview the video lecture about pn junction. In the beginning, Alice just linearly watches the video. Although she knows that there are some differences between P material and N material in pn junction, the underlying difference between them explained from this lecture is quite new and interesting for her. Thus, she marks the video part as "important" and "pn junction" so that she can rewatch it later. After a while, Alice feels that this lecture seems somewhat trivial. She decides to skim it to see if there is something she probably shouldn't miss. Soon, she finds out that there is a tutorial and starts watching the corresponding video part. She gets stuck by one key step of the tutorial exercise. She decides to tag it as "difficult" and then go back to digest it once she finishes watching the whole tutorial. She still has several video parts which are tagged as "difficult" and "confusing" when she finishes watching the whole video. As she has other assignments to do, she plans to discuss those parts with her classmates and instructor next week in the classroom.

3.3 Overview of Quick Tagging Interface

A screenshot of the quick tagging interface can be seen in Figure 3.1. In the middle, the large picture is the main viewer, similar to most video players. Clicking on the main viewer allows the user to toggle between playing and pausing of the video. This same functionality can also be found in YouTube's player as well as various video viewing applications found in the mobile space. On the bottom is a toolbar that houses video controls, allowing the user to play or pause the video and view the current playing time. Users

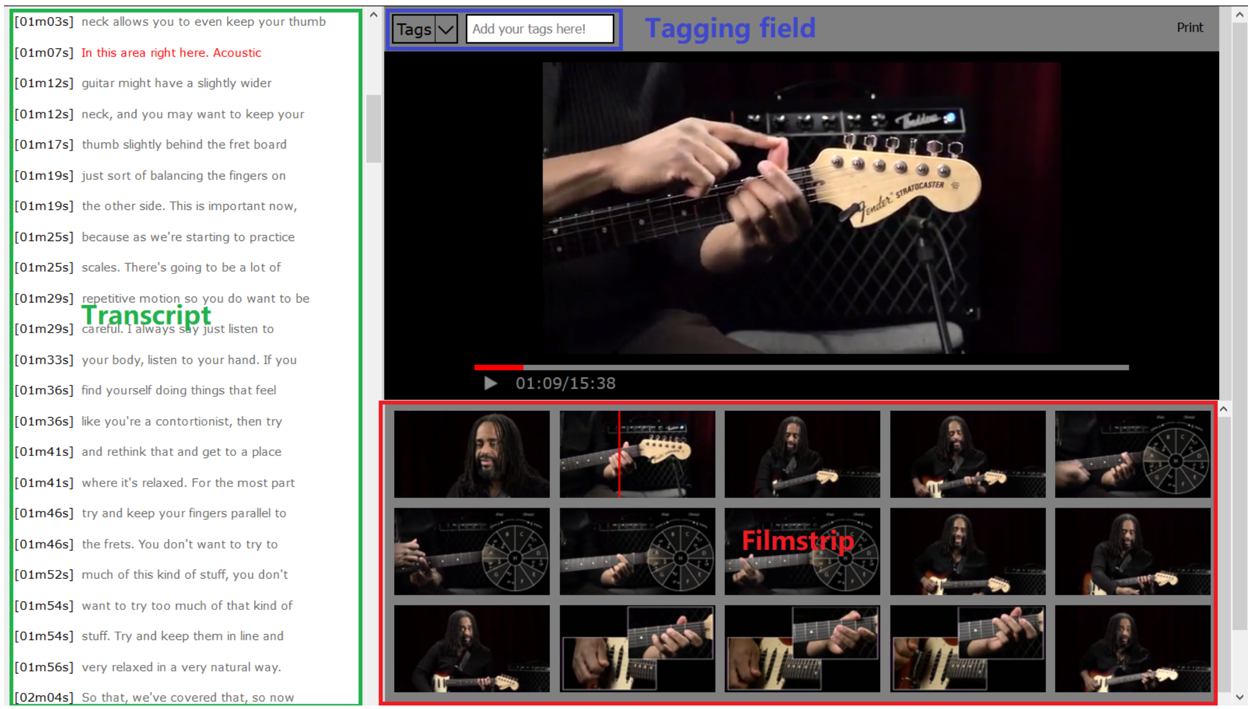

Figure 3.1: Overview of the quick tagging interface. Here, we see the transcript (green), the filmstrip (red), and a tagging field (blue).

can also skip forward or backward five seconds by using the arrow keys.

The rest of the interface, such as the filmstrip (red, Section 3.3.3) and the transcript (green, Section 3.3.4) will be described in detail below. We first describe our pre-defined tags.

3.3.1 Pre-defined Tags

We focus on two classifications of our pre-defined tags in learning contexts to help recall video content, namely reaction tags which are used to express viewer feelings or opinions, and topics which describe video content. As reaction tags are subjective, one of our assumptions is that the negative and positive valence and the intensity (e.g. like and love) of such tag words will be biased under different learning contexts. To maintain low cognitive load and save interface space, the optimal default tag number with highest preference and least tagging efforts should also be considered. Topics, on the other hand, are objective and strongly related to course subjects and content. In Chapter 4, we discuss how we got the most used tags and the optimal number of tags to display in our quick tagging interface as well as in other design implications.

In Figure 3.3, we can see that there are ten pre-defined tags which are used in our lab study in Chapter 5. The five reaction tags are chosen from study results in Chapter 4 according to the metrics of high use frequency and word valence and intensity under learning contexts. The five topic tags are extracted from video content and a field is available for users to input their own tags. We have made this decision with the hope that our participants can focus on the quick tagging mechanism instead of experiencing a lack of availability while inputting tags. To uphold consistency throughout the entire interface, the tagging field will be used in playhead tagging, transcript tagging, and filmstrip tagging.

3.3.2 State Diagram of Quick Tagging Mode

As shown in Figure 3.2, there are two modes in our quick tagging interface, namely tagging and untagging mode. In tagging mode, there are three tagging methods: playhead tagging, transcript tagging, and filmstrip tagging. For playhead tagging, users can use one hot key (Ctrl + z) to perform one-click tagging. Here we used hot key to achieve the one-click interaction on desktop. Our study was not focused on a particular hot key in the tests so it was selected for convenience. There was no undo mechanism implemented. Thus, as there was only one hot key that the participants had to learn with,

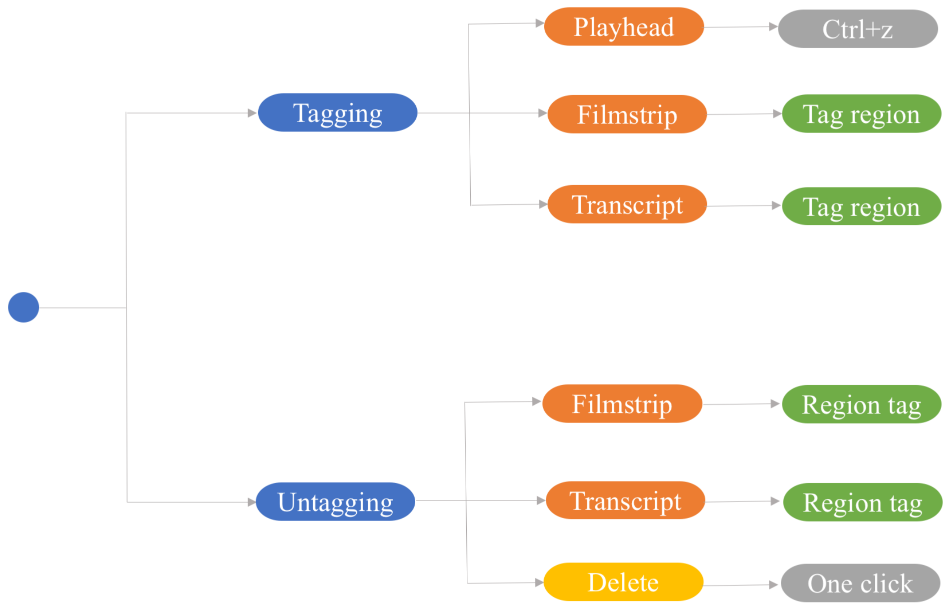

Figure 3.2: Tree state diagram of three levels of quick tagging mode. From left to right, the first level has two modes: tagging and untagging. The second level shows two branches of three tagging each, for the allowed tagging actions in corresponding mode. The third level shows two tagging orders and other ways of interaction.

overloading will not be an issue. For transcript tagging, users choose tags first and select video content (tag region) in tagging mode. A similar order applies to filmstrip tagging in tagging mode. However, in untagging mode, there is one tagging deletion action and two tagging methods, called transcript tagging and filmstrip tagging. Contrary to the order in tagging mode, both transcript tagging and filmstrip tagging need users to select video content first and then choose tags (region tag) in untagging mode. In general, Noun-Verb order is best for users who prefer to use the same tags to tag all video content. Alternatively, the Verb-Noun order works well for users who frequently change tags. In untagging mode, users can delete their taggings with one-click. Further descriptions will be provided below.

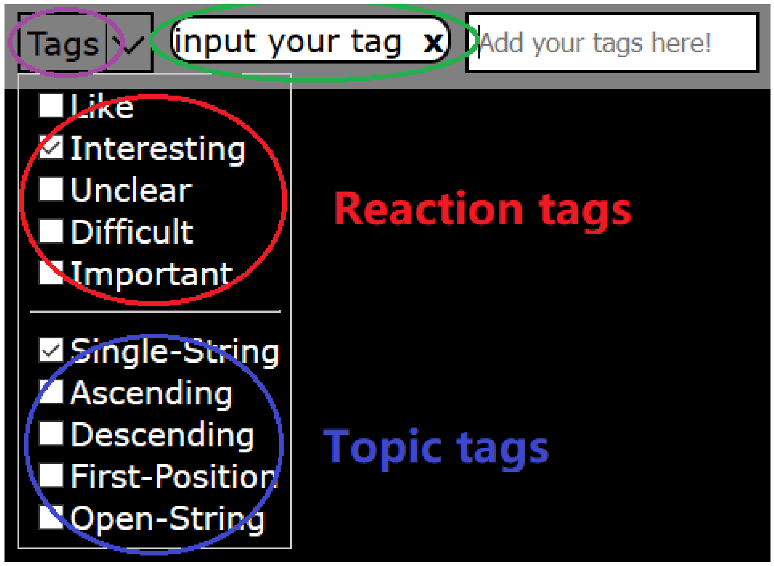

Figure 3.3: Tagging field with user input tag (green) and ten pre-defined tags comprised of five reaction tags (red) and five topics (blue). The tagging mode ("tagging" and "untagging") is switched by a toggle button (purple). In this figure, "tagging mode" is displayed.

3.3.3 Filmstrip

The filmstrip is the interface element bound in red in Figure 3.1. The filmstrip allows users to seek to any part of the video, and provides users with a preview to allow better seeking accuracy, as well as timestamps to help judge the location of the potential seek.

The filmstrip is viewed as a set of thumbnails from the video arranged side by side, each representing a portion of the video. As the width of the filmstrip represents the entire length of the video, each n thumbnail represents 1/n of the video. Granularity should depend on video length. Moving the cursor over top changes the corresponding thumbnail to show the frame represented by the horizontal location of the cursor and corresponding timestamp. The initial image is the first frame of the represented interval. The red bar acts like a playhead.

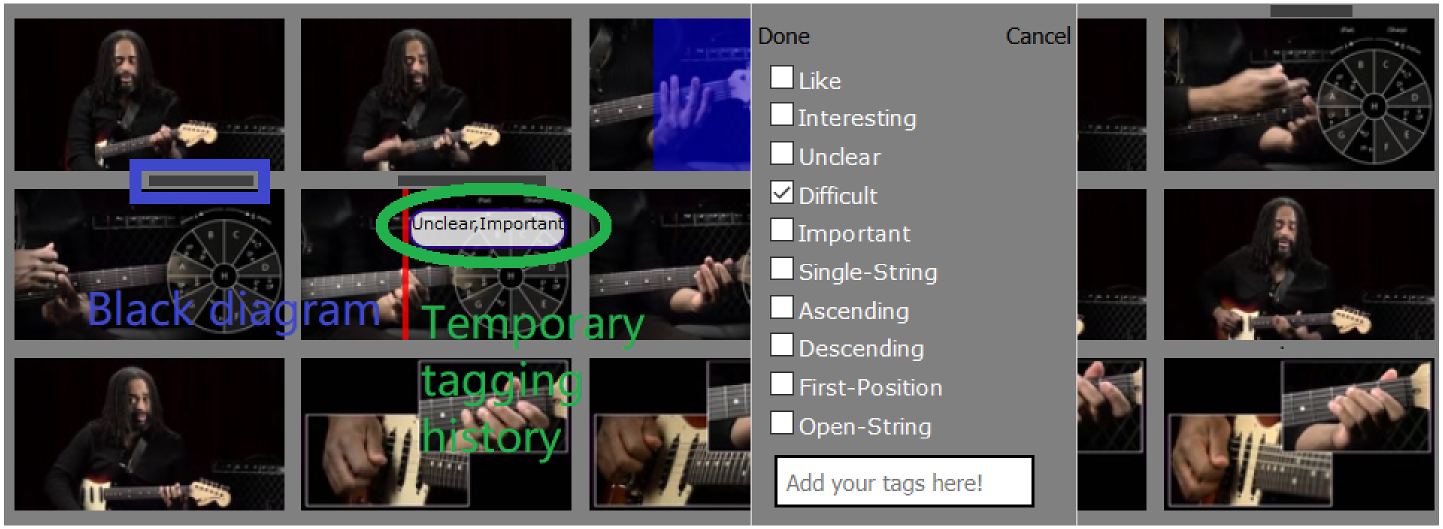

Figure 3.4: The filmstrip splits into multiple rows, each representing a portion of the video. Black diagram (blue) displayed on the top of each thumbnail represents the tagging history.

Applying Tags on Filmstrip

There are two ways to apply tags on a filmstrip. One way is called playhead tagging where users can use one hot key (Ctrl+z) to add tags. To perform playhead tagging, users need to be in tagging mode and select pre-defined tags or input a tag from the tagging field, shown in blue on the top of Figure 3.1. The video part will be the current spoken sentence. More specifically, the tagged video interval will be the sentence around the playhead time where the tagging action occurs. The video length of the sentence being spoken is known from the audio transcript.

Another method is called filmstrip tagging, shown in Figure 3.4. There are two orders available here. One is to select thumbnails first and then choose tags (or input one from tagging field) to apply tags on filmstrip (region tag). The other order is to choose tags (or input one from the tagging field) and then select thumbnails to apply tags on the filmstrip (tag region). Figure 3.4 shows the tag region order in untagging mode. Region tag order can also be achieved in tagging mode, by choosing tags from the drop-down menus in Figure 3.3 and then selecting thumbnails to apply tags.

Tags Empowering Navigation on Filmstrip

The black diagrams (highlighted in blue) in Figure 3.4 will be displayed on the top side of the thumbnail to indicate the tags made in the corresponding video part. If there are multiple tagged video parts for each thumbnail, the opacity of overlapped black diagrams will be darker. The user can click the black diagram to navigate the video. To provide more text information, the corresponding tags will pop up when hovering on the black diagram. Correspondingly, the temporary tagging history (highlighted in green) will show up beside the red bar when the tagged video part starts, and disappears after it ends.

3.3.4 Transcript

The transcript, shown on the left in Figure 3.1 in green contains a transcript of everything said in the video. This provides an overview of the video in textual form, allowing users to quickly scan through the spoken content of the video. As the video plays through, the sentence being currently spoken turns red and acts like a playhead. On the left of each caption is a timestamp, which is clickable to allow users to jump around the video.

Applying Tags on Transcript

Like the filmstrip, the user can perform both playhead tagging and transcript tagging to apply tags on a transcript. Similarly, the two orders can be used for transcript tagging in two tagging modes. The bottom of Figure 3.5 in blue shows the order to select texts first and choose tags (or input a tag) to apply tags on the transcript during untagging mode.

Tags Empowering Navigation on Transcript

After the user tags a video, the tagging history will be shown in the transcript. The tagging history, shown on the top of Figure 3.5 in red, is comprised of a box to circle around the selected texts, an arrow, and a vertical tag bubble. The user can click this tag bubble, causing the player to seek the video part from the start. This navigation function can be quite helpful when users browse through the transcript and tagging history, especially for longer videos. In addition, users can refer to the tag history and use the timestamp to navigate.

3.4 Design Strategies

In general, quick tagging here refers to a function that provides an easy and fast way for users to mark course video parts with predefined semantic keywords. More specifically, users can easily and quickly choose tags and apply them to course video content. The "quickness" of this design is amplified by three aspects:

Predefined tags: In a study by Sen [33], the author claims that 68% of non-taggers in the study did not tag because they could not think of any tags. We suppose this might also be an issue for users who are non-English native speakers. In light of this, we provide pre-existing tags for users to choose from, hoping to ease this dilemma. These tag words are extracted from comments and annotations of course videos in YouTube and CLAS and are noted as the most commonly-used by people when describing video content. These pre-existing tags can even save time for English native speakers. In IBM's Efficient Video Annotation(EVA) tool [38], all annotations apply terms from a small controlled-term vocabulary, and no free text annotations are allowed. This promotes consistency, simplicity, and speed of annotation.

Video part selection: In [8] and [20], the appropriate granularity of video segmentation is an important issue to consider for quick video annotation and tagging. According to the interview results from [8], the most useful granularity of segmentation in general is the shot level, namely a video segment with in and out points. Since video represents a special case of a segmented continuous variable [8], and video parts are used as the navigation points inside educational videos [15], we decided to choose video part as our tagging application in this paper. Here, the video part shares the same granularity of segmentation with the shot level. However, unlike Videotater [8], precision of segmentation is not our priority. So, in order to make the video part selection process fast and simple, we use an auto-selection method.

More specifically, once users find video points they want to mark, they can easily generate video parts with one click. According to [10], the familiar interaction of highlighting text was preferred, which inspired us to use text selection and filmstrip selection as two other ways for selecting video parts.

Applying tags: The evaluation results of [38] tell us that the tagging task can be more efficient when tagging multiple frames for only one concept at a time. This conclusion was based on the assumption that users can clearly recognize and easily locate the corresponding frames. Drawing upon Baudisch's work that used a painting metaphor to rate large numbers of objects [3], the order of selecting items and methods was categorized as noun-verb and verb-noun. Considering the diversity of use behaviors in tagging videos, one of our design decisions was to integrate the order of video part selection, and include tag application with one click. We also decided to support both noun-verb and verb-noun orders. In other words, users are able to select text or filmstrip first, and then apply tags, and vice versa.

3.5 Summary

In this chapter, we introduced our quick tagging interface, which is composed of the main viewer, transcript, filmstrip and tagging field. The "quick" is enhanced from three perspectives: pre-defined tags, video part selection, and tag application. Our video part selection methods can be widely applied to various video content (such as videos with strong transcription but poor visualization, good visualization, and so on). We have supported two orders for applying tags: tag region and region tag. Region tag is best for users who frequently change tags, while tag region works well for users who prefer to use the same tags consistently. We then used two modes to help users switch between the two orders, as well as other interaction methods (like tagging deletion). In Chapter 4, we will discuss how we choose pre-defined tags for educational videos, and we will determine the optimal number of tags in our interface. Then, in Chapter 5, we will compare our quick tagging mechanism with hand-written notes in a controlled lab study. In the notes-taking condition, the plain video interface will comprise of a main viewer, transcript, and filmstrip. In the quick tagging condition, these three components (main viewer, transcript, and filmstrip) together with the tagging field, will be used.

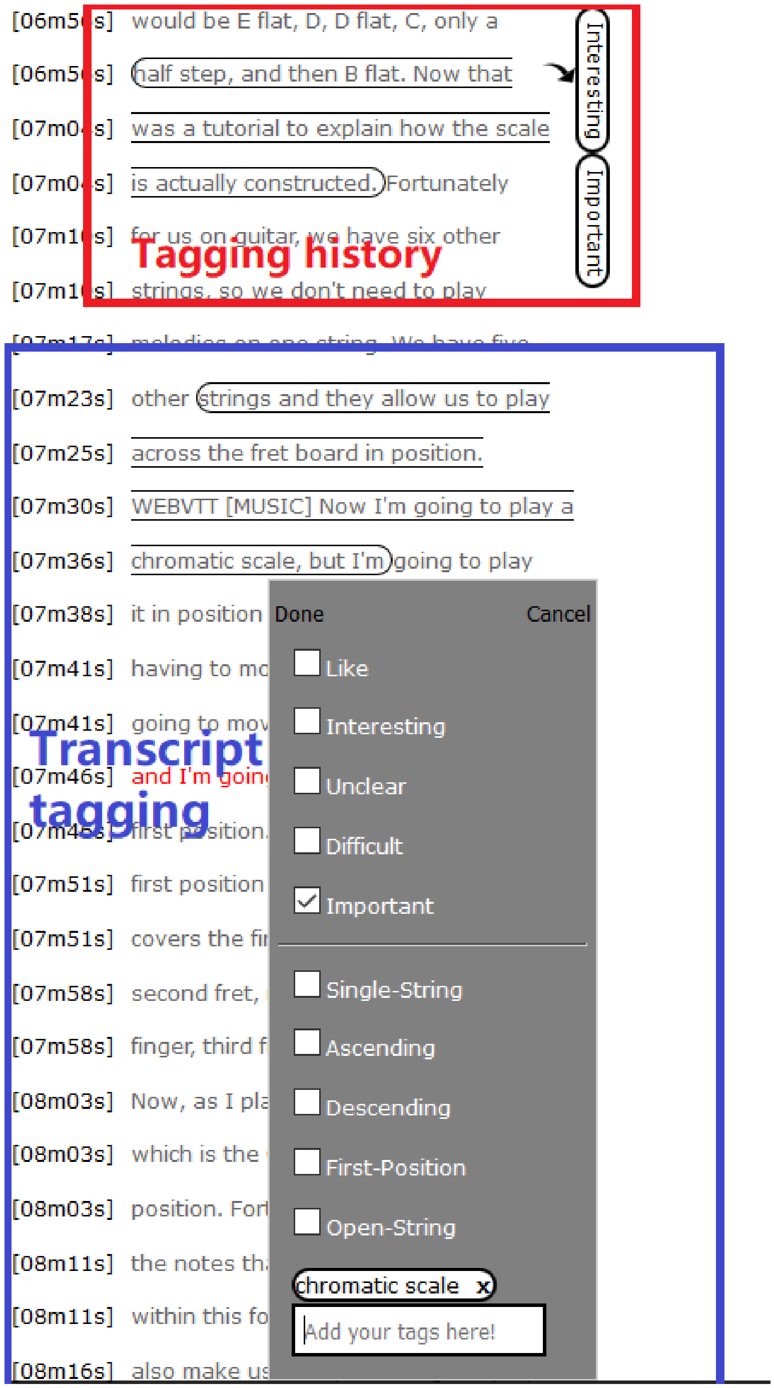

Figure 3.5: The transcript shows the tagging history and a tagging field.

Chapter 4

Background Study

In this chapter, we discuss our background study conducted to choose predefined tags under learning contexts. This study was divided into two parts. First, we evaluated a taxonomy method for tag words from two datasets and collected the most widely-used tags in three categories. Second, an online survey was conducted to validate the aggregated results from the evaluation and develop the design implications for our quick tagging interface.

4.1 Datasets Analysis Methodology and Evaluation

In this section, we first introduce the taxonomy methodology for analyzing video comments. We then describe the two datasets separately, from the YouTube educational video channel and the Collaborative Learning Annotation System (CLAS) [2], a video platform for learning. Considering availability, we were only able to access YouTube and CLAS datasets. In our content analysis, we used only philosophy subject on YouTube but many other subjects on CLAS. Notably, we found that there were larger amounts of comments available in philosophy courses than other subjects' on YouTube. In our definition of three categories of tags, the general tags applied to all subjects and only the topic tags were dependent on subjects themselves. In light of this, we believe this kind of distribution will not bias our results. Finally, we provide the analysis procedure and final results in this section.

4.1.1 A Taxonomy Method of Tag Words

For analysis logic, we want to find the most common words of users to express their feelings and opinions while watching videos, as well as how they describe the video content. We manually perform colour-coded word taxonomy for the extracted words from video comments. There are three main steps for this process.

[2] http://clas.sites.olt.ubc.ca/

Admiration/Awe	Desperation	Happiness	Lust
Amusement	Disappointment	Hatred	Pleasure/Enjoyment
Anger	Disgust	Hope	Pride
Anxiety	Dissatisfaction	Humility	Relaxation/Serenity
Being touched	Envy	Interest/Enthusiasm	Relief
Boredom	Fear	Irritation	Sadness
Compassion	Feeling	Jealousy	Shame
Contempt	Gratitude	Joy	Surprise
Contentment	Guilt	Longing	Tension/Stress

Table 4.1: Affect categories by Klaus. There are 36 categories in total.

Choosing Appropriate Tag Words

We call this first step, the “word cleaning step”. We filter out many irrelevant words, mainly prepositions, articles, and personal pronouns, as well as some nouns such as “video”, and verbs such as “think” or “have”. We do not consider words of estimative probability, such as “probably” or “certainly”, for these have strong context dependency in a sentence. Here our research aims for one-word selections. Emotion words such as “wow” or “haha” are not semantic words, so we also remove them from this study. Although many typos are encountered, such as “ndesirable”, it is quite obvious to recognize them, notably for words with ‘n’ in the comment text.

Categorization

Second, we categorize the cleaned words into three categories: opinion words, content descriptive words, and subject feature words. Klaus [32] claims that words to express emotions for an object are called “discrete emotion words”. In his paper, people’s affections are classified into 36 categories, and a list of pertinent words in each corresponding category is provided. All of these categories are shown in Table 4.1. We found that most of the

Category	Function	Scope
Opinion words	Express opinions and watching feelings	General
Content descriptive words	Describe video content and express opinions	General
Subject feature words	Describe video content	Context-based

Table 4.2: Definition of each category from function and scope. Here words in first two categories can be applied to general course videos. Words in last two categories both can be applied to describe video content, but the latter one is course context-based.

words in Klaus' work can be applied to express opinions in the context of tagging educational video content, as well as expressing feelings while watching videos. Therefore, in our work to classify opinion words, we directly refer to these 36 affection categories. For our other two categories, there are no existing lists we can refer. We just have a general sense that content descriptive words will mainly be composed of subjectively adjective words and subject feature words mainly composed of objective nouns and some adjective words only used in courses of specific subject. The border line for some adjective words between the two categories is still vague. Analyzing a broad range of subjects to validate the words in content descriptive category and subject feature category can be one way to solve this issue. More details are described in our evaluation section. Definition of each category is shown in Table 4.2.

Word Frequency Analysis

We get a big picture after this second step. But there is still a great amount of words in each category, and we find that many words have different forms but share the same meaning. For example, "interesting" has other forms such as "interest", "interested" and "interests". What's more, words like "happy", "glad", and "joyful" share similar meaning. So we decide to combine words with different forms and similar meanings into one word set, and choose the highest frequency word as representative word of the corresponding set. We apply this inductive process to the three categories. For opinion words, we decide which words are synonyms by referring to listed word stems of each affect category from Klaus 4.1. As for the other two categories, three researchers in Human Computer Interaction (HCI) work together to validate the results of synonyms. The main researcher (myself) categorized the first version. Then another researcher validated and modified the results of the first version. Finally, the third researcher did the last round of validation and modification.

4.1.2 Datasets Description

We analyzed video comments from two datasets: YouTube and CLAS. There are two goals of this evaluation. The first is to analyze a broad range of subjects to validate words in the content descriptive category and subject feature category. More specifically, if words are categorized as subject feature words, but also appear in other subjects, we categorize them as content descriptive words. The second goal is to output an educational video tag form sorted by frequency, and collect the most used words in each category for the next online study. The reasons for why we use comment words as keyword tags is two-fold. Firstly, there are no keywords for tagging educational videos available in current literature that we are aware of. Second, it is clear that many comments in educational videos include reactions, opinions and topics directly related to the video content itself.

YouTube Dataset

In the YouTube dataset, we analyzed 4870 comments from philosophy courses; 39 philosophy course videos are used, 14 from the CrashCourse channel[3] and 25 from The School of Life[4] channel. We scraped all comments from these videos using the YouTube API, and counted how many times each word was used. Finally, we input all words with appropriate frequency to Microsoft Excel for analysis.

CLAS Dataset

The CLAS dataset consisted of data from students in Music, Math, Library and Information Studies, Political Science, and Art History. The amount of usage in each class differed based on the amount of influence the instructor had on encouraging students to use the system: students were either told that viewing and posting counted for participation marks, or viewing

[3]https://www.youtube.com/user/crashcourse
[4]https://www.youtube.com/user/schooloflifechannel

Subjects	**Video #**	**Comment #**	**Annotation #**	**Posting Type**	**Year Level**
Music	3	46	321	Encouraged	1st year(1)
Math	18	154	3522	Encouraged	1st year(2)
Library	33	67	794	Encouraged	Graduate(1)
Political	25	4	28	Voluntary	1st year(2), 3rd year(1)
Art	98	0	420	Voluntary	1st year(1), 2nd year(1), 3rd year(2)

Table 4.3: Details of CLAS data. For example, in the first year music course there are 3 videos with 46 comments and 321 annotations, with encouraged posting.

Affect_category	Pertinent_words	Representative_word
Interest/Enthusiasm	curious (15), enthusiastic (1), interesting (176), entertaining (8), involved (13), engaging (4), disengaged (1), attractive (9), appealling (6), compelling (2)	interesting

Table 4.4: An example process of choosing representative words from a word set. Here we take the Interest/Enthusiasm affection category as an example word set. This comprises a group of pertinent words with similar meanings. “Interesting” has the highest frequency in the word set, and is thus chosen as the representative word.

and posting was completely voluntary. Within each course, the posting comprised of two parts: aggregate annotation content (a comment for a moment of video) and aggregate comment content (a comment on the whole video). Details about metrics (such as annotation type, posting type, and year level) of each course subject are shown in Table 4.3.

4.1.3 Procedure

We first used the three-step word taxonomy method (Word Cleaning, Categorization, and Word Frequency Analysis) to analyze the comments on philosophy courses from YouTube. These word taxonomy steps are common in literature and current applications. We then normalized the representative words in the opinion words category. We did not normalize words in the other two categories, because further validation first needs to be conducted.

We needed to clarify that each category was comprised of word sets. A group of words with similar meaning made up one word set, and the word with highest frequency was chosen as the representative word in that word set. An example of this process is shown in Table 4.4. The resulting normalized use percentage of each representative word was in fact the result of the total word count in each word set, divided by the total word count in each word category. The normalized results were then compared with results from the CLAS dataset. As there was initially only one subject in this YouTube dataset, the validation was completed after finishing the analysis of another five subjects in CLAS; there are five course subjects in the CLAS data. For each subject, we followed the same analysis procedures as we performed with our YouTube data. However, as the data metrics among these five subjects are slightly different, our analysis procedures needed to adapt to the corresponding data features. In other words, the three-step word taxonomy method was applied at subject level in the YouTube dataset, while applied at metric level within each subject in the CLAS dataset.

Our analysis and comparison was completed within each subject in this step. For the courses in music, math, and library, there is only one year level for both annotations and comments, so We first completed a word taxonomy separately for annotations and comments within each subject, and then compared the results. For the political science courses, there are two year level for both annotations and comments. Here, we first performed word taxonomy separately for annotations and comments within each year level, and then compared the results between the two annotation types and year levels. For art history courses, we see three year levels with just annotations, so we first performed word taxonomy on annotations within each year level.

	Music	**Math**	**Art history**	**Political science**	**Library**	**Philosophy**
true	✓	✓	✓		✓	✓
new	✓	✓	✓		✓	✓
wrong	✓	✓			✓	✓
limited		✓	✓		✓	✓
pretty	✓	✓			✓	✓

Table 4.5: Five words were moved after cross-validation study. Check mark means the word originally existed in this subject.

We then compared the results across the three year levels. As shown in Figure 4.1, the top five words in annotations are almost the same with ones in comments which respectively applied to the three word categories. Other results of this intermediate step are attached in the Appendix C.

From the results of our last step, we concluded that the data metrics such as year level and annotation type had no significant influence on taxonomy results. Based on this finding, we integrated results into three categories for each subject separately, and performed analysis and comparison across subjects. We compared words in the opinion, content descriptive, and subject feature categories, respectively, across the five subjects. We then manually verified the generality of words in both opinion and content descriptive categories, and the context-based property of words in the subject feature category. Thus, we integrated words in the opinion category across five subjects into an aggregated opinion category for CLAS. We applied the same procedure for words in the content descriptive category. Following our YouTube routine, we only normalized words in the opinion category. Finally, we obtained opinion words and content descriptive words from the YouTube and CLAS data respectively.

We also had subject feature words from six course subjects (5 from CLAS). For feature words that exist in more than three of the course subjects, we reclassified them as content descriptive words. After this validation step, we finalized words in both the content descriptive and subject feature categories. We then separately normalized words in the content descriptive category in each dataset. Comparing normalized opinion words from YouTube

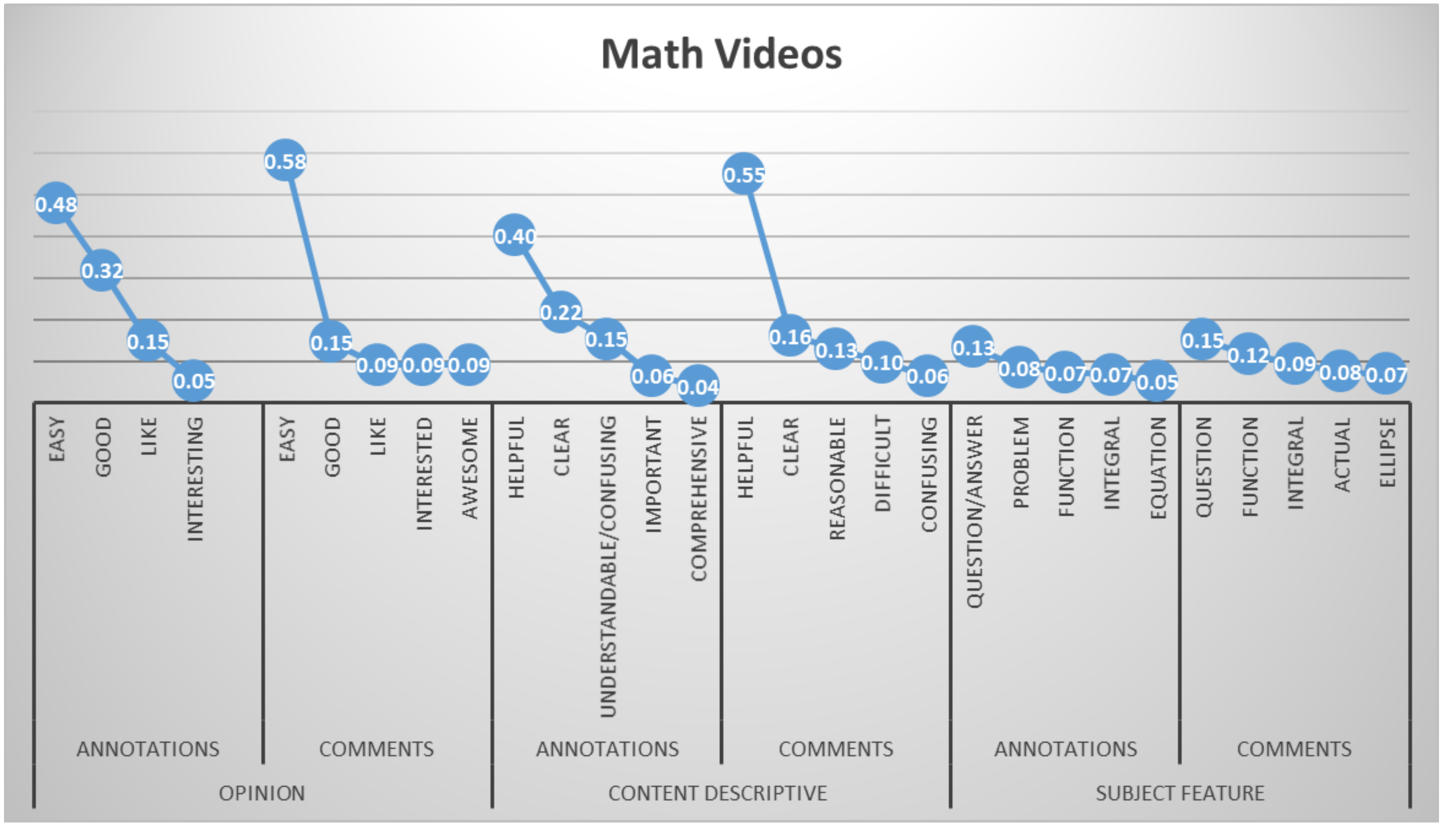

Figure 4.1: Comparison results between annotations and comments in math courses from CLAS. The blue bubbles show normalized percentages. The comparison was run among three categories: opinion words, content descriptive words and subject feature words.

Subjects	**Feature Words**
Music	dynamic, confident, rhythmic, loud, rush
Math	question, function, integral, answer, actual
Art	ritual, authentic, social, modern, aesthetics
Political	deal, constitutional, works, vulnerability, unanimous
Library	instructional, topic, question, handout, active
Philosophy	factual, stupid, absurd, popular, logical

Table 4.6: Top five collected subject feature words in six subjects, respectively.

with ones in CLAS by use percentage, we formed the aggregated opinion words category, and applied the same procedure to words in the content descriptive category. The aggregation process is attached in our Appendix D. Lastly, we employed a form that contains one aggregated opinion words category, one aggregated content descriptive words category, and six subject feature categories. We selected the top collected words in the two general categories to perform our next study.

4.1.4 Results

For CLAS data, results within each course subject showed that in all music, math and library courses, the top used words in three categories are almost the same between annotations and comments. In other words, annotation types do not influence the top used words across our three categories. This conclusion also applies to year levels. For courses with voluntary postings, words mainly belonged to the subject feature category, and were few in our other two categories.

In Table 4.5, five words moved from the subject feature to content descriptive category and were listed. These listed words had very high use percentages. We can see that “true” and “new” existed in five subjects, while the other three words existed in four subjects. We also observe that political science feature words do not contain any of these five words. We might attribute this to voluntary postings and fewer quantities of annotations may contribute to this, but we also see that art history courses have voluntary postings. In any case, all five words were contained in the feature words category. Thus, from another perspective, we suppose that the subject feature words in political science are completely distinct from ones in other subjects, which enhances our decision to classify subject feature words and

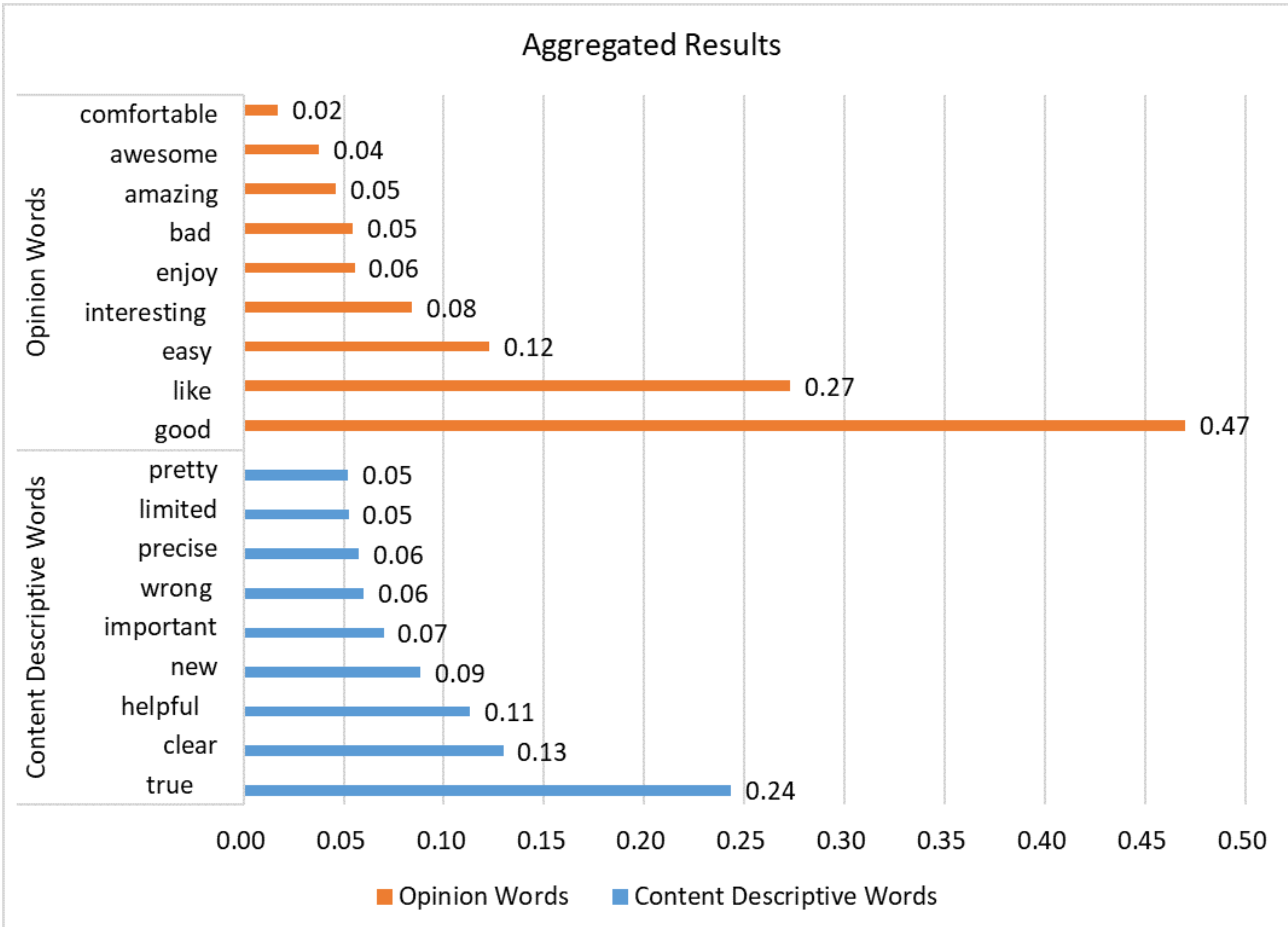

Figure 4.2: Aggregated results from Normalized YouTube and CLAS data.

verify our definition for this category.

In Table 4.6, the top five feature words in six subjects are given. We can see that the words in each subject are completely different and are context-based. The word "question" exists in math and library with different frequencies, but it does not exist in more than three subjects, so we still classify it as a subject feature word. We can see that these results are made up of nouns such as "function" and "topic", and adjectives such as "rhythmic" and "absurd". Further, the distribution of nouns and adjectives is dependent on subject. Math mainly comprises of nouns, while philosophy and music are made up of adjectives.

In Figure 4.2, the top nine words in opinion and content descriptive categories were given, respectively. As mentioned previously, these words will be used in our next study. We observe that most words are positive and that few words are negative, such as "bad", "limited" and "wrong". For negative words, "bad" and "wrong" are from the YouTube dataset. In general, more negative words were collected in the YouTube datasets than in CLAS. This conclusion can be proved in our Appendix D. We also observed that words like "good", "like" and "true" had significantly higher use percentages than others' in the list.

4.2 Validation of Top Collected Tags

An online survey was conducted in order to assess the usability of the collected tags and to determine an optimal tag number. With different numbers (3,5,7,9) of given default tags, participants were asked to tag five educational video clips and were encouraged to input their own tags.

4.2.1 Apparatus

Five educational video clips from YouTube were embedded in the survey. The video clips were roughly 20 seconds in length and were completely different from videos in the first study of Section 4.1.2. Five subjects were covered: music, art history, engineering, math, and information management.

4.2.2 Participants

Thirty-five subjects participated in the study and we compensated them with free coffee. They were randomly invited to take part in the study by our four researchers in Kaiser building in UBC. Two laptops and two iPads

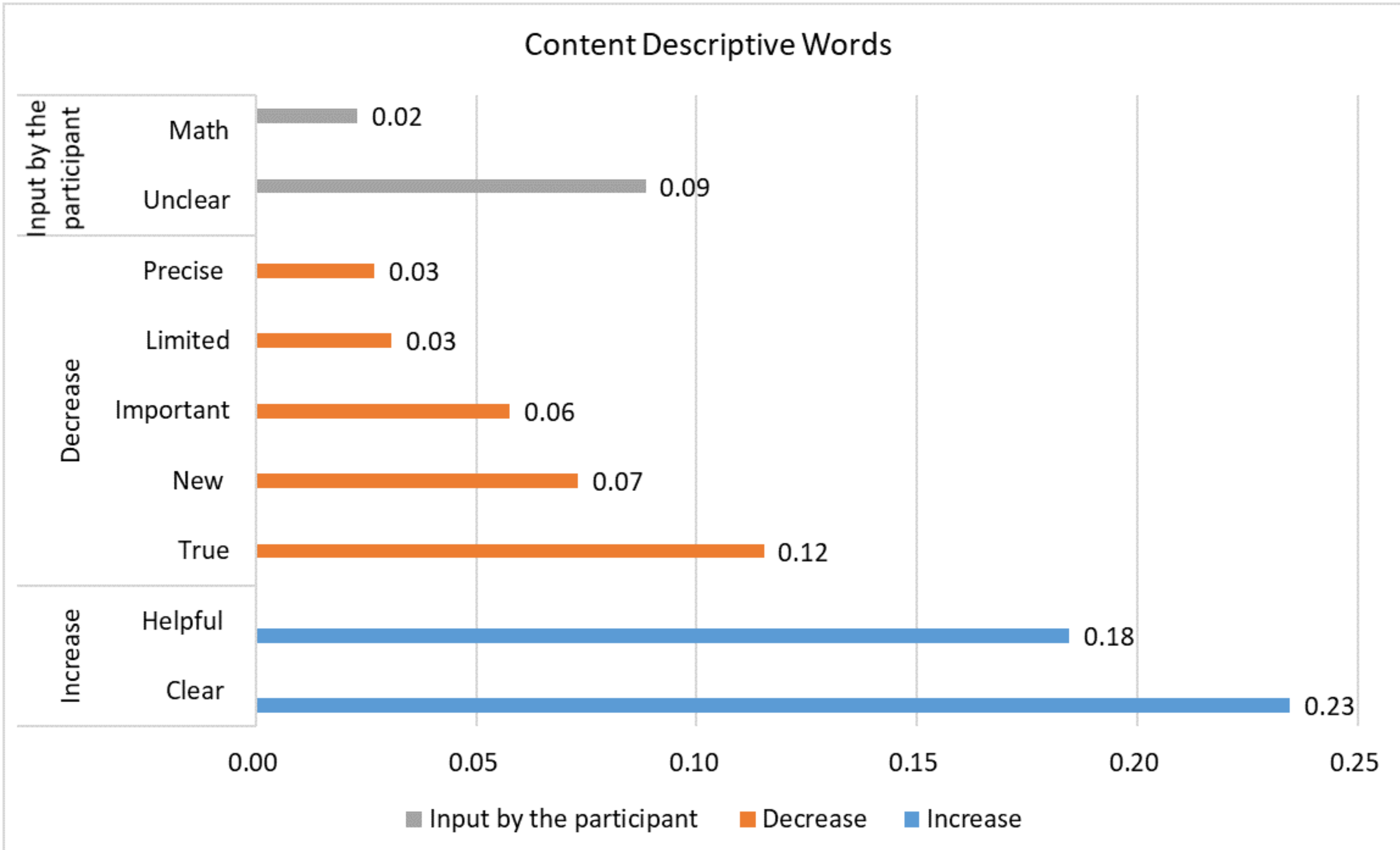

Figure 4.3: Content descriptive word results in the survey compared with normalization results in YouTube and CLAS data. There are three groups: words input by the participant (gray), percentage decrease (orange) and percentage increase (blue).

were separately put on four tables. Each device was used to run one tag group (4 groups in total). Two researchers were responsible for instructing the participants to do online survey with the available devices. As the four devices were not occupied at the same time, the two researchers were also responsible for assigning participants to corresponding device to make sure that all participants were randomly and evenly divided into four between groups. Another two researchers were responsible for inviting students to take part in the study by introducing the study purpose, time length, and reimbursement. As it took participants at most five minutes to finish the study, many students who were in their class break or waiting for their friends were willing to take part and got free coffee. All thirty-five participants took part in the online survey within one day. Three participants did not complete the survey and were excluded from the analysis. The other thirty-two participants (8 female, 24 male) were all university students (aged 19-40, with one male student less than 19). The thirty-two students (14 English native speakers and 18 non-English native speakers) were from four faculties: Applied Science (N = 24), Graduate and Postdoctoral Studies (N = 5), Land and Food Systems (N = 2), and Pharmaceutical Sciences (N = 1). Nine students had no experience taking online courses taught with videos.

4.2.3 Procedure

A 4x2 factorial study was designed for this survey. Four between groups were respectively assigned 3, 5, 7 and 9 default tags. We followed the routine for designing Likert and semantic differential scales [30] where 3, 5, 7 and 9 are often used. Tags in both opinion word and content descriptive words were provided to each group.

Participants were asked to watch an educational video clip first and then respectively select tags that could best express their feelings while watching and their descriptions of the video content. After finishing and tagging five video clips, participants were asked to choose default numbers of tags they would like to have. Finally, questions about tagging task difficulty and tag usefulness were asked.

As the online survey itself is within group, we made four survey versions to control the four levels of default tag numbers. Each participant was randomly assigned to each version. To avoid the order effects of given tags, all the tags were randomly presented for each video clip. The video clips were named “video1-5” to avoid any bias from the real names of the videos.

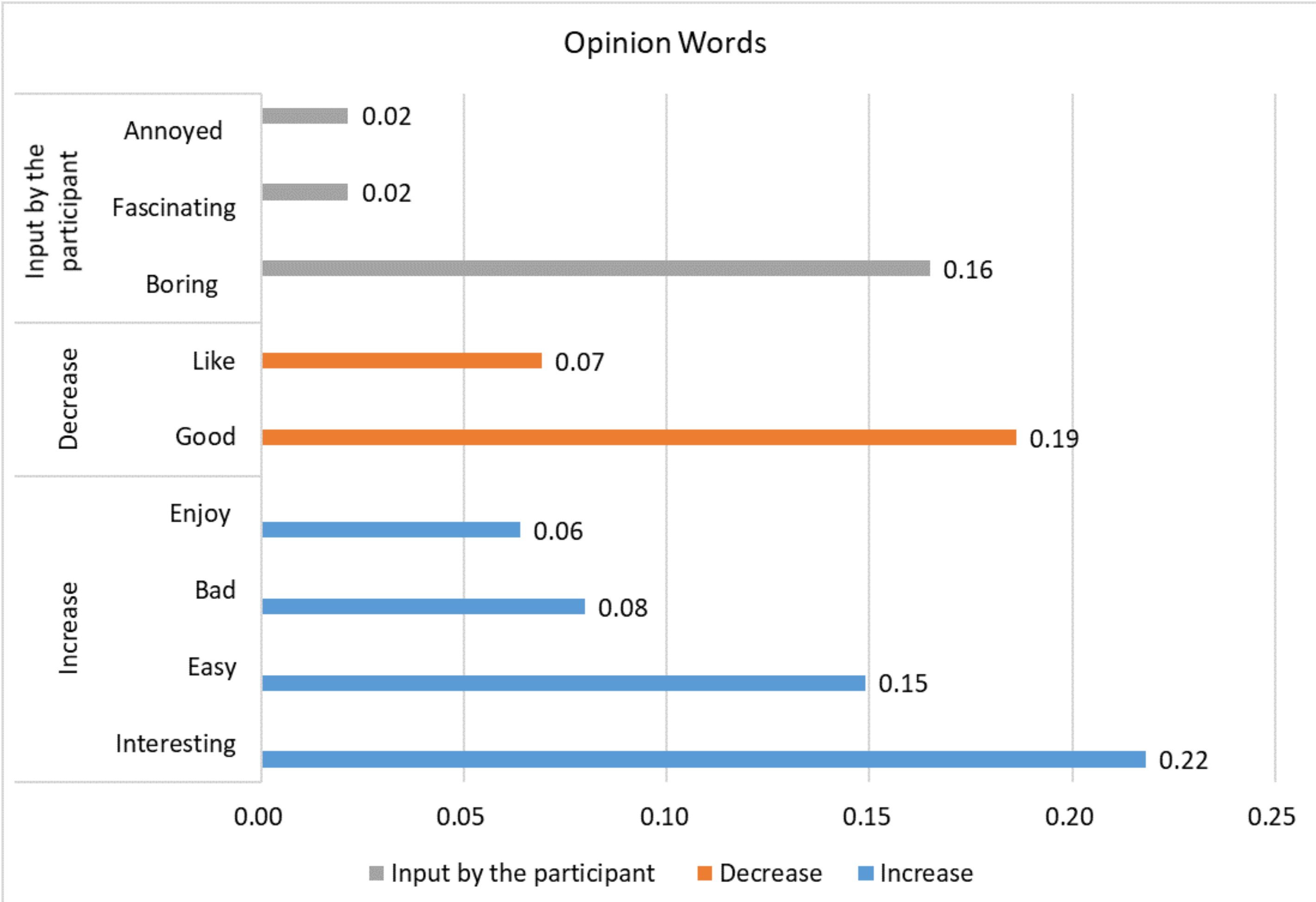

Figure 4.4: Opinion word results in the survey compared with normalization results in YouTube and CLAS data. There are three groups: words input by the participant (gray), percentage decrease (orange), and percentage increase (blue).

4.2.4 Results

We used data and theory triangulation to analyze the results and answer the two research questions below.

Q1: Are the given words really useful and representative of reaction range for tagging?

Survey questions were asked about whether the provided tags were useful to describe video content and express watching feelings and opinions respectively, and participants were asked to rate on a five Likert scale. A Kruskal-Wallis H test was then run to determine if there were differences in usefulness to describe video content among four groups of participants given different numbers of tag words: 3 words(n=8), 5 words(n=7), 7 words(n=8), and 9 words(n=7). Here, the n amounts refer to group size numbers. Values are mean ranks unless otherwise stated.

Distributions of usefulness to describe video content were not similar for all groups, as assessed by visual inspection of a boxplot. The usefulness to describe video content increased from 9 tags (10.43), to 5 tags(14.71), to 3 tags(16.25), to 7 tags(19.88), but the differences were not statistically significant, $\chi^2(3) = 5.227$, p = .156.

The same test was run to determine if there were differences in usefulness to express watching feelings and opinions among the four groups given different numbers of tag words: 3 words(n=8), 5 words(n=8),7 words(n=7), and 9 words(n=7) groups. The mean rank of usefulness to express feelings was different between groups, and was statistically significantly: $\chi^2(3) = 9.369$, p = .025. Both p-values agree. According to [6], the asymptotic p-value (2-sided test) is considered good enough when there were five or more participants in each group. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. This post hoc analysis revealed statistically significant differences in usefulness to express watching feelings and opinions between 5 tags (10.38) and 7 tags (23.14) (p = .016), but not between 9 tags (14.43) or any other group combination.

The Group with 3 words, namely "clear", "helpful" and "true" thought given words were quite useful to describe video content though the difference was not significant. In Figure 4.5, we can see that the sum of the three words' use percentages is above 50% among the four groups. This suggests that the given three words are quite useful. The group with 7 words had the highest rating on usefulness to describe video content, which may be caused by the provided negative word, "wrong". We can see that the 3 words group and 5 words group were all given positive words in Figure 4.5. Our

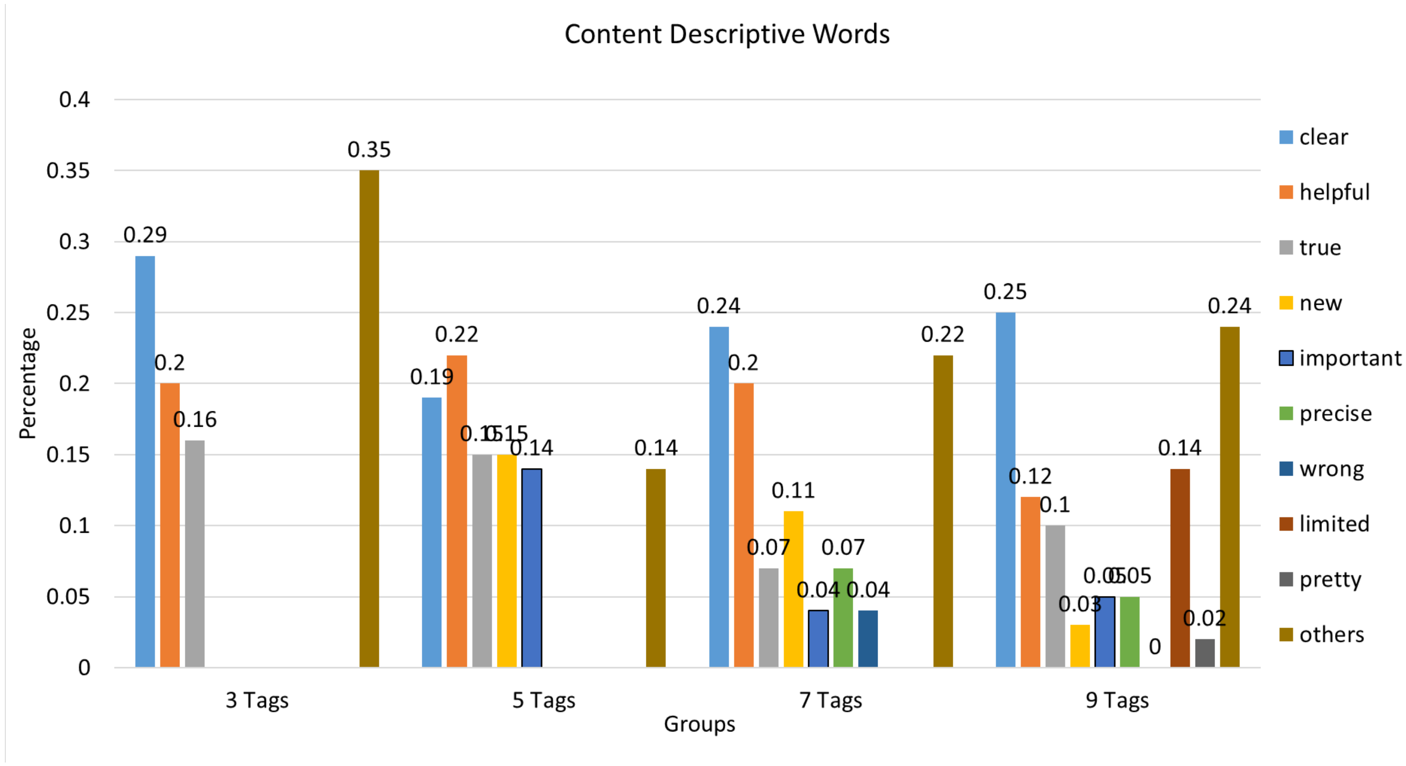

Figure 4.5: Content Descriptive word use percentage in each group. The listed nine words are default words. “Others” means words input by the participant.

argument can be enhanced by results in Figure 4.3 where we see that the negative word "unclear" was input by users and is in the top four. Although the 9 word group was also given the negative words "wrong" and "limited", the group had the lowest rating. This can be explained by the fact that participants thought only some of the given words were useful, such as the negative words and the top three words.

This analysis can also be applied to the results of rating usefulness to express watching feelings and opinions. We can see that the negative word "bad" was given in both the 7 words and 9 words groups from Figure 4.6. Thus, the group with 7 tags rated significantly higher than the 5 tag group. The usefulness of negative tags can also be seen by Figure 4.4 where the negative word "boring" was input by users and places in the top three.

Q2: Which is the optimal default tag number?

Participants were asked which number was preferred to display default words in a paper prototype. 1, 3, 5 and 7 were provided as default selections, but manual input was also supported. Of the 32 participants recruited to the study, 4 preferred to have 1 word, 7 preferred to have 3 words, 14 preferred to have 5 words, 5 preferred to have 7 words, and 2 preferred to have 4 words. A chi-square goodness-of-fit test was conducted to determine whether the tag numbers mentioned are equally preferable. The minimum expected frequency was 6.4. The chi-square goodness-of-fit test indicated that the default selections were not equally preferable by the participants ($\chi^2(4) = 13.313$, $p = .010$), with 3 words and 5 words preferred by students.

Participants were also asked to score the amount of effort it took them to add tags for the video clips in a five Likert scale. A Kruskal-Wallis H test was then run to determine if there were effort differences in adding tags among the four participant groups given different numbers of tag words: 3 words(n=8), 5 words(n=8),7 words(n=8), and 9 words(n=7) groups. Values are mean ranks unless otherwise stated. The distributions of effort to add tags were not similar for all groups, as assessed by visual inspection of a boxplot. The effort to add tags increased from 5 tags (13.75), to 3 tags (14.06), to 9 tags (16.21), to 7 tags (20.00), but the differences were not statistically significant: $\chi^2(3) = 2.707$, $p = .439$. The 5 tag group had the lowest effort scores, which may suggest that they felt most comfortable with 5 tags.

In both Figure 4.5 and Figure 4.6, we can see that the 5 tag group also had the lowest percentage (0.14 and 0.17) of input words compared with the other three groups. To some degree, this suggests that 5 default words should be the most suitable number, given the various trade-offs found in the analysis. Additionally, The five default tags should include both positive

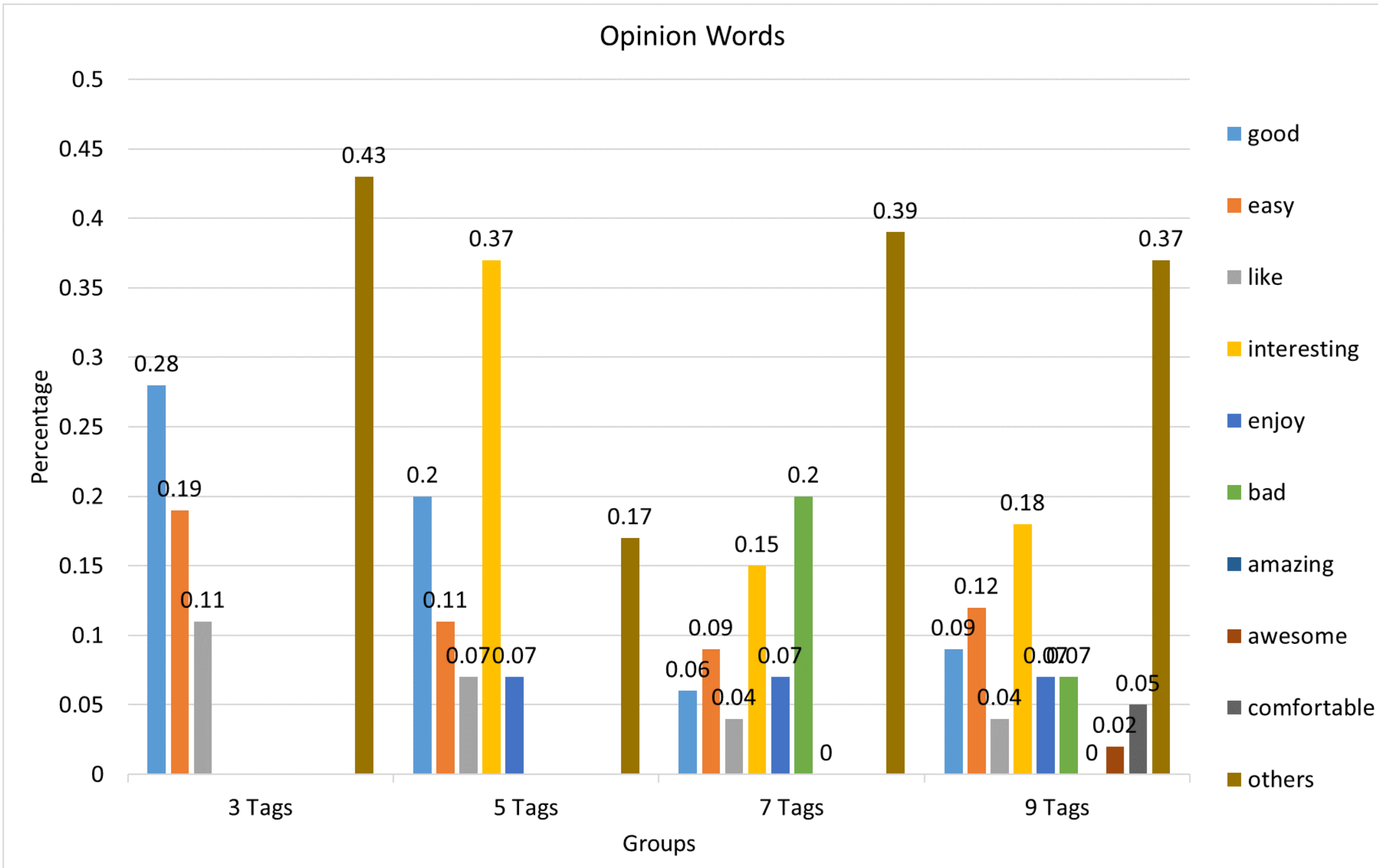

Figure 4.6: Opinion word use percentage in each group. The listed nine words are default words. “Others” means words input by the participant.

and negative valence words.

4.3 Discussion

A tagging interface for students to efficiently mark up video content can benefit from the restricted tag vocabulary. These tags can also play an important role as an alternative mechanism to commenting, using less screen real estate. When choosing pre-defined tags for online video study, the learning contexts should be considered (such as personal vs. public, mid-term, quiz, etc.). In addition, the factors of tag reaction range and word intensity should be coupled with each other. More specifically, how to expand the reaction range of tagging in a more fine-grained way can also be explored in future research, besides simply including positive and negative tags. Another future direction is in associating learning action with reaction tags. For example, “interesting” may lead to sharing or re-watching the tagged video content, while “bad” may cause others to skip the video content entirely. In addition, “important” may lead to more reviewing of the video content in the future.

As for our tag taxonomy method, we only consider the frequency factor. Because our work focused more on design of the quick tagging interface, we chose not to consider other factors. How to choose representative tags could certainly be more comprehensive, such as distinguishing “love” and “like”, etc. Specifically, more factors such as association and intensity can be taken into account.

According to Hick’s law [9], the more time will be taken to manipulate the interface the more options there are in the interface. Survey design for tagging tasks in each group would be a future direction for calculating the Tagging Tasks Finished Time Difference (TTFTD) as a function of design.

4.4 Summary

This chapter introduced a tag taxonomy method, which extracted the most commonly used tags from video comments. We evaluated this method separately for two data sets and aggregated the results from normalized YouTube and CLAS data. The top collected tags were further validated by an online survey. According to our results, five is the optimal tag number for a one-click tagging interface. Additionally, a tag reaction range should cover both positive and negative polarity. Learning contexts should also be considered when choosing the pre-defined tags for educational video interfaces. Based on these design implications, we recommend five reaction tags: “like”, “interesting”, “unclear”, “difficult”, and “important” for a quiz preparation learning context, which will be used in our lab study prototype in Chapter 5. Other directions can be explored, such as how to extend the current taxonomy method, associating learning action with reaction tags, and utilizing the Hick’s law.

Chapter 5

Lab Study

In this chapter, we discuss the experiment to evaluate our interface. This evaluation was performed following pilot studies to polish and receive initial feedback from users about the usage of a quick tagging mechanism for video learning tasks. Here the pilot studies helped us make design decisions for the quick tagging interaction and evaluation methods of our controlled study. In previous iterations, our quick tagging interface focused on supporting playhead tagging and only one tagging order but pilot studies have shown that varied methods are needed to support users' tagging actions efficiently and that two tagging orders are required to satisfy different user behaviors. Thus, we introduce our quick tagging mechanism integrating playhead tagging, transcript tagging, and filmstrip tagging, with pre-defined tag words.

5.1 Experiment

A user study was carried out to evaluate the design and performance of our interface and the usefulness of using the quick tagging mechanism to help users recall video content. We developed an evaluation protocol to encourage users to use our quick tagging method while keeping the experiment short and maintaining a bias free evaluation. Likewise, for comparing with current practices in quick tagging, we needed to ensure that our interface mimicked current approaches as well as their logical extensions to provide a fair comparison. Using our protocol, we investigated whether integrating the quick tagging mechanism into the current video player would make recalling video content more efficient in a learning context. We conducted a comparative user study, comparing the performance of participants using a quick tagging mechanism to bookmark video content against watching the plain video interface and taking notes on paper to finish a quiz corresponding to the course video.

5.1.1 Apparatus

The experimental interface was implemented using HTML5 and JavaScript. The study was performed on a 13.3 inch Acer S7 Laptop, with a screen resolution of 1920x1080 pixels. Participants used a regular Microsoft mouse to manipulate the contents of the screen. The web browser was Firefox (version 53.0.2, zoomed to 100%). We also used a screencast tool named Camtasia Studio 8 to record participants' interface operations, as shown in Figure 3.1.

5.1.2 Participants

Fifteen volunteers, 6 males and 9 females, participated in this experiment. They were compensated $20 Starbucks gift card for their time. Participants ranged from ages 19 to 40. Each participant worked on this task individually. Twelve participants had taken a video course previously, either from UBC or from other on-line services. Three participants had never taken any video course. Participants were from diverse fields (7 Applied Science, 1 Business, 1 Arts, 2 Education, 1 Graduate and Postdoctoral Studies, 1 Land and Food Systems, and 2 Science). None of the participants had a strong background in guitar, which was the topic of the video in the experiment. Participants were aware they could opt out of the experiment at any time, which implied that they did not have to complete tasks as prescribed. Two participants gave up passing the quiz from the second half of the video. Therefore, their corresponding data was removed from our quantitative analysis. According to our evaluation protocol, if participants could not pass the quiz (4 out 5 questions right) within 20 minutes, they would be noted as failing the quiz. This experiment was time-limited, as we needed to control user study length to avoid bias due to tiredness from lengthy learning sessions. One participant failed both of the quizzes. Another participant failed one quiz from the first half of the video, thus, their corresponding data was removed from our quantitative analysis.

5.1.3 Design and Procedure

We separated participants into two groups. One group took handwritten notes during the first half of the video and switched to the quick tagging interface for the second half (paper-first condition). The other group did the opposite (tagging-first condition). The thirty-minute video was chosen from an introductory guitar course from Coursera[5]. We chose the guitar video for three reasons:

- **Tagging motivation and preference:** We piloted three kinds of video lessons: Guitar, History and Chemistry. Participants showed interest in history and music videos, but disliked chemistry. According to our piloting, we found that when people felt interested in the video topic, this would influence their motivation to tag video content while learning. In addition, the history video had a good transcript, so participants preferred tagging just on the transcript. The other two videos had strong transcripts and filmstrips, so participants liked to tag both. However, the guitar lesson also had many visual demonstrations to complement the audio.

- **Quiz completion:** We modified the quiz from Coursera. Each quiz is comprised of five multiple choice questions, and we observed that participants tended to complete the history quiz by simply searching for key words in the video transcript. The chemistry quiz also involved heavy calculations, which may have taken significant amounts of time to learn, slowing down completion time. The guitar quiz struck a good balance between these two issues. As we discussed in our related work in Section 2.2, the guitar video content and quiz were best able to simulate the learning process from memory retention to analysis.

- **Video length:** We learned from our piloting that short videos were not able to motivate participants to use our tagging functions for learning in the lab experiment. As a result, participants mainly relied on their memory instead of using the tagging functions we provided to finish the quiz, so we were not able to verify the usefulness of our tagging interface. In other words, longer videos can reduce this learning effect to some degree, which the guitar video (15 minutes) can accommodate for our study.

Each participant was asked to play the role of a distance learning student. Specifically, they were told that their instructor assigned a video for them to watch and a quiz for them to finish. Using the quick tagging interface, participants were encouraged to use the tagging functions to bookmark video content. We hoped that participants would develop their own use and learning strategies for the tagging interface. For comparison, participants

[5] https://www.coursera.org/learn/guitar/home/week/4

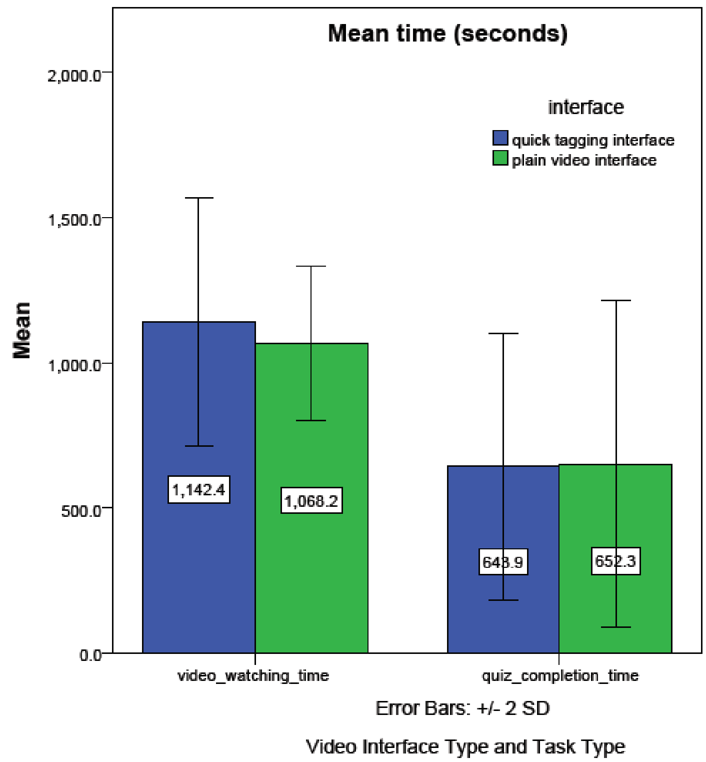

Figure 5.1: Descriptive Statistics for quick tagging vs plain video interface. Mean times are shown for each task and interface. Numbers are in seconds.

also were instructed to use a plain video interface that did not support tagging, where they were allowed to take notes on paper. The guitar video was equally divided into two halves for each participant to learn with the tagging interface and with the plain video interface. The order of each interface was randomly assigned to reduce order effect. Participants were told that they could use the video and tags or written notes as a cheat sheet to help them finish the quiz later. The time for participants to learn and watch each video was recorded and a maximum of 30 minutes was given for learning and watching the 15-minute video.

The participants were asked to complete the quiz as soon as possible. They were required to get four out of the five questions right to pass the quiz. The time taken to pass each quiz was recorded, and the timing begun when the user opened the quiz. We then checked the answers when participants

Question	Median	p value
Overall usefulness of recall	4	.003
Overall ease of use	4	.071
Overall efficiency of use	4	.065
Overall helpfulness for learning tasks	4	.003
Pre-defined tags are helpful to recall	4	.048
Pre-defined tags are useful to mark down	4	.005
Liking transcript tagging	5	.001
Liking filmstrip tagging	3	.782
Liking playhead tagging	3	.317
Tagging efforts on transcript	2	.034
Tagging efforts on filmstrip	2	.031
Tagging efforts by playhead	2	.003

Table 5.1: The aggregated results of our questionnaire, with the median score (Likert Scale, 1 to 5) and p value (significant difference from neutral reaction). Here, 1 represents strongly disagree and 5 represents strongly agree. Participants' overall reaction to our quick tagging mechanism was highly positive.

finished the quiz and the timing only stopped when they passed the quiz. We indicated their wrong answers, and allowed them to continue to attempt the quiz until they passed, with a time restriction of 20 minutes.

The experiment proceeded as follows:

1. The participants started with a pre-study questionnaire to fill in their demographic information and usage history of on-line course videos.

2. The evaluation started with a proficiency test for each technique. Participants were shown how to use each of the interface elements, and most importantly, how to tag in both tagging mode and untagging mode. They were also shown how to navigate the video content through tagging history, transcript and filmstrip, and how to access tagged video parts with the visual cues (black diagram) on the filmstrip. The video used in this proficiency test was a popular course from a UBC professor [6] on Field Effect Transistors (FET), which was 30-minutes long. After this, they were required to finish each instruction of the proficiency test independently. After passing the test, they were allowed to proceed to the formal study section.

[6] https://www.youtube.com/watch?v=SjeK1nkiFvI&t=141s

3. After the proficiency test, a trial began by asking participants to watch and learn from the first half of a given video (the guitar video). Participants were able to make a cheat sheet using the learning method. Once completed, they were asked to finish a quiz related to the video content.

4. Once completing this first quiz, participants were advanced to the second half of the video and were instructed to use the alternate learning method. Once completed, they were asked to finish a quiz related to the video content. Upon completion of the whole video, participants were given a questionnaire to fill out asking for their reactions to the quick tagging interface and comments.

5. Finally, participants were given time to experiment with the quick tagging interface in a less structured environment, and a follow-up interview was conducted to further explore the tagging design space as well as other usage scenarios.

Each experiment took between 1.5 hours and 2 hours.

5.2 Results and Discussion

We compared times spent finishing the quiz, when using either the quick tagging method to mark down video content or by hand-written notes with the plain video interface. The time to finish the quiz by using the quick tagging interface (M=643.91, SD = 229.74) and by plain video interface (M=652.27, SD=280.60) was shown in Figure 5.1. The reason why we used descriptive data analysis instead of a paired t-test was that the data was not normally distributed as well as having large standard distribution which violated the requirements for a robust paired t-test. As we mentioned previously, four participants did not pass the quiz and their data were excluded from descriptive data analysis. We also found that bookmarking video content with the quick tagging interface took participants 10% longer on average than when taking notes on paper.

There are many factors which can influence the time participants take to pass a quiz, which we consider here. As discussed in our related work in Section 2.2, our quiz needed participants to apply their memory-related learning skills and analytical learning skills, which simulates real-life learning scenarios. In other words, when participants tag video content, this does not mean they fully understand the content or are freely able to apply the knowledge. In contrast, written notes may in fact be a better method allowing participants to understand the content, to some degree. In addition, our quiz is comprised of multiple choice questions, where participants had the opportunity to randomly choose an answer. In other words, they could have simply used trial and error quickly until answering enough questions correctly. All of the above factors may contribute to the insignificant difference in finishing times between the two conditions. As seen in our observations, participants could also review the tagged content and then move on to finish the quiz. However, when participants took notes on paper, they would directly go to the quiz when the video ended. In addition, some participants commented that some design decisions in the quick tagging interface were inefficient. More details will be discussed in Section 5.2.2. These findings may explain why it took participants more time to use the quick tagging interface than when taking notes.

The questionnaire results shown in Table 5.1 demonstrate the positive reaction of participants to our quick tagging interface and mechanism. They especially agreed that the quick tagging interface was significantly useful to recall video content (Median=4, p=.003) and that the quick tagging mechanism was significantly helpful to finish learning tasks (Median=4, p=.003). Participants generally thought the interface was easy to use (Median=4, p=.071) and efficient (Median=4, p=.065), though the difference here were not significant. Here we provide a quotation from P13:

> "I think the video tagging function is really helpful for students especially during lectures, as professors tend to speak fast and long, and students can only capture a point or two and jot them down in a hurry. When they re-read their notes, there's missing pieces in connecting the dots and it doesn't make sense." [P13]

In the later sections of this thesis, we further present and discuss our findings about how the participants used our quick tagging interface to facilitate their learning process while watching the video. Specifically, we will focus on:

1. Pre-defined tags (such as the number and variety of the tags, whether they were satisfying or not, the most used tags for learning, polarity of the tags, participant's need to input their own personal tags, and the number of tags they apply at one time).

2. Quick tagging modes and interaction (such as preference, suggestions, use patterns, etc.).

Reaction	Frequency	Topics	Frequency
Important	168	First-Position	13
Interesting	34	Half-Step	9
Difficult	7	Descending	8
Unclear	4	Ascending	7
Like	1	Open-String	6
		Key	3
		Sequence	2
		Single-String	0

Table 5.2: Pre-defined tags in our interface and aggregated use frequency by 14 participants. The total number of use frequency of general tags was 214, and was 48 for that of topic tags.

3. Further improvements for the tagging interface (design and visualization, suggestions on adding new functions, aesthetics).

We used a triangulation method to analyze our data from an observation of participants' use patterns, their questionnaire reactions, and their time taken to finish the tasks.

5.2.1 Pre-defined Tags

The pre-defined tags are one of the most important parts of our proposed quick tagging mechanism. We will elaborate on the present results and discussion with three aspects: use frequency and usefulness of pre-defined tags, satisfaction with pre-defined tag quantity and diversity, and requirements for input tags.

Are the default tags used frequently and are they useful enough for learning?

In terms of the pre-defined tags, we provided five reaction tags to express watching opinions and five topic tags related to video content to help participants bookmark video content, as shown in Table 5.2. We can see that the total use frequency number of reaction tags (214) was around 4.5 times more often than topic tags (48). As for reaction tags, words with positive meaning, such as "important" and "interesting", were used substantially more often than negative words like "difficult" and "unclear". Three participants (P2, P3 and P9) used both "important" and "interesting". Four participants (P11, P12, P13 and P14) indicated that they used "important" as if they were highlighting. Specifically, P12 mentioned that the study context was for quiz so "important" was useful enough but for other learning contexts, more tags would be used. Three participants (P1, P3 and P11) used negative tags in the study. Another three participants (P12, P13 and P14) indicated that they would like to use negative tags such as "difficult" and "unclear" if they learned their own courses so that they can review the corresponding video content later. However, one participant (P9) would never use "Unclear" and "Difficult", and explained that P9 would try to work out difficult content until it became clear. Most participants thought "Like" was improper in the quiz preparation context. Participants did not input any reaction tag in the study, so we could not find an appropriate alternative to "Like" for a time-constrained personalized learning context. As "Like" had high use frequency in Chapter 4, we still recommend to use "Like" as default reaction tag for general collaborative learning contexts.

When asked about whether the given tags were satisfactory in terms of quantity and variety, both P3 and P6 thought they were good enough. P3 also mentioned that the given tags helped him focus on video learning and had lower cognitive load than thinking about his own tags. Notably, P12 only used reaction tags. Others did use topic tags, but the ones provided were quite limited, resulting in five participants (P2, P5, P8, P9 and P15) inputting their own. Correspondingly, we observed that all the tags input by participants were topic words according to our experiment logs.

In general, participants agreed that the topic tags would help them gain an overview of the video content. But when it comes to tagging video content, some participants would use reaction tags while others would input their own tags instead. No participants used the tag "Single-String". On the one hand, it could be explained that some participants used other tags to bookmark related content. Alternatively, we observed that some participants struggled on the quiz question related to the single string topic, meaning that these participants did not pay attention to the corresponding content while they watched the guitar video. Thus, both P3 and P10 suggested that it would be helpful to display topics tags physically related to corresponding video content instead of displaying them in a menu which loses spatial information, especially when people were not familiar with the video topic.

As shown in Table 5.1, participants agreed that the pre-defined tags were significantly helpful for recall (Median=4, $p=.048$) and useful to mark down video content (Median=4, $p=.005$). According to our interviews, most participants liked the provided tags. However, P7 thought the tags contained not enough information and preferred to take notes. To some degree, this

Multiple tags	#	Multiple tags	#
Interesting,Important	11	Ascending,flats	2
Important,movable fingering	2	Descending,flats	2
Difficult,Important	1	Descending,alternate picking	1
Interesting,Important,First-Position	1	Ascending,Descending,Open-String	1
Important,second position	1	Ascending,sharps	1
Important,First-Position,open string	1	chromatic scale,left hand relax	1
Important,octaves	1	Ascending,Descending	1
Important,first position	1	First-Position,Open-String, g chromatic good for fingering	1
Important,Chromatic scale	1	First-Position,Open-String, two octaves apart	1
Interesting,Key	1		
Interesting,movable fingering	1		
Interesting,alternate picking	1		
Interesting,legato	1		

Table 5.3: Multiple tags and aggregated use frequency(#) by 6 participants. The number of tagging applications from 14 participants was 262 in total. Multiple tagging was around 13% (35/262) of the total tagging.

explains why P7 did not tag at all in the experiment. Similarly, P5 took many notes on paper and preferred to input their own tags.

Is there a need for using multiple tags? If yes, what should they look like?

From Table 5.3, we can see that six participants used more than one tag in the experiment. According to our interviews, four participants indicated that they would only use one tag. Six participants would like to use more than one. However, one participant added that they would not use more than two. Indeed, P4 used only one tag in the experiment, but indicated that more than one would be used for more complex and longer videos.

In Table 5.3, there were three types of multiple tags, namely reaction tags only (12), reaction and topic tags (12), and topic tags only (11). These were almost equally distributed. In our interview, P14 mentioned that two reaction and one topic tag would be used, and P15 preferred to input multiple tags.

Discussion

From our findings, we can see that the pre-defined tags are indeed helpful for recalling video content. There are some design implications from using both reaction tags and topic tags in future tagging interfaces. For reaction tags, people tended to use the pre-defined tags without inputting their own. Based on this finding, transforming the reaction tags to emoji or icons could be a promising future research direction. Emoji or icons can save design space on the video interface and are less cluttered than text. In addition, both positive and negative tags should be considered to satisfy users' needs.

From the perspective of topic tags, there are two main functions: an overview of video content and bookmarking video parts. According to our findings, people need more variety and larger numbers of topic tags than what was provided. Thus, a crowd-sourced tag cloud could be a solution to provide an overview of video content. Dynamically chaptering the topics can also be a promising solution to closely relate topics to video content. Learners who are unfamiliar with the video content can potentially benefit from this solution.

Support for inputting a user's own tags is necessary for a complete tagging mechanism. There are a few arguments to support this conclusion. First, users would like to input their own phrases or words together with pre-defined tags. Second, some users have strong opinions on typing their own topics. Finally, this can be sourced by a tag cloud. Many factors such as saving tag input by users and how to better support tagging actions in the case of multiple tags can be considered in future design.

5.2.2 Quick Tagging Modes and Interaction

As described in Chapter 3, there are two modes: tagging and untagging. The quick tagging interaction involves video part selection followed by applying tags, and vice versa. Three methods are supported to perform quick tagging, namely transcript tagging, filmstrip tagging, and playhead tagging. For spontaneous tagging, the process of video part selection can be achieved with one hot key interaction.

As for transcript tagging shown on the bottom in Figure 3.5 in blue, the interaction process involves selecting texts in the transcript first and then choosing tags/inputting tags, and vice versa. Similarly, filmstrip tagging (Figure 3.4) can be achieved in two different orders: selecting filmstrip first and then choosing tags/inputting tags, or vice versa. The results and discussion of this section are elaborated by the preference, suggestions, and use patterns of the quick tagging mode and interaction.

Comparison of three quick tagging methods

Results derived from the questionnaire showed that participants significantly liked transcript tagging (Median=5, $p=.001$) and felt neutral about filmstrip tagging (Median=3, $p=.782$) and playhead tagging (Median=3, $p=.317$). However, it took a significantly small amount of effort for them to tag the video content in all three ways: transcript (Median=2, $p=.034$), filmstrip (Median=2, $p=.031$), and playhead (Median=2, $p=.003$).

According to our interviews, one participant liked playhead tagging best which helped mark down video content while focusing on watching. However, 12 participants preferred transcript tagging for the following three reasons. First, participants were generally not familiar with the video topic and reading text helped them learn faster. Second, some participants were used to reading and highlighting on PDF, therefore they naturally selected text on transcript and correspondingly applied tags. Third, participants liked to select exact words or sentences when they performed tagging, which explains why they disliked the filmstrip. More specifically, they thought the visualization of the filmstrip was quite similar in each frame and that selecting with the filmstrip gave them information that was too general and inaccurate. However, four participants mentioned that it would be very handy to tag charts or graphs on the filmstrip once they were familiar with the video topic.

Suggestions on improving the quick tagging modes and interaction

Four participants felt that it was confusing to have two modes. They suggested to instead use a context menu with hot keys to commit all the actions: add tags, edit and delete tags/tagging, as well as undo. Here is a quotation from P10 below.

> "It could be useful to add functionality to the right button of the mouse, to open a menu box for creating or deleting tags." [P10]

The other eleven participants thought the two modes were acceptable, but six of them suggested that the tagging mode should only allow adding tags while the untagging mode was best for deleting or editing tags/tagging. As for the two different orders of quick tagging: selecting video content first or applying tags first, three participants liked both orders. Four participants used the same tags frequently so they preferred the second order, while three other participants frequently changed tags so they preferred the first order.

In our design, the drop-down menu does not automatically close after one tag is chosen, to support the selection of multiple tags. Three participants complained that this design was annoying and two participants thought changing tags in the drop-down menu was time-consuming. Keyboard shortcuts were suggested to efficiently change tags. Below are two representative quotations to clarify the two issues above:

> "Not have to close the tag/untag drop-down menu before the tag is saved because sometimes the user forgets to do that and the tag does not refresh." [P2]

> "It would be nice to be able to change the tags more quickly. It takes time away from watching the video when you have to go back and forth to select different tags." [P8]

Use patterns of the quick tagging modes and interaction

According to our observations, almost all participants(13) chose tags first and then selected video content. Six of them only used transcript tagging. Four of them used both transcript tagging and playhead tagging. Two of them only used playhead tagging. Three of them occasionally used filmstrip tagging. One typical scenario was seen when there was only music and no transcript, where participants would select on the filmstrip. Here, two participants would first select texts on the transcript and apply tags. Six participants edited their tagging/tags by adding a new tag, changing tags, as well as modifying selected texts in the transcript. In addition, one participant performed editing after finishing the video, while others edited while watching. Although we did not specifically provide an editing mode in our interface, participants developed their own strategies here. For example, five participants moved to the untagging mode and manipulated the tagging application (trimmed the selected texts, interacted with the tag bubbles, deleted the tagging application, and selected again to add new tags). However, one participant tagged twice on the same video content to add new tags in tagging mode. In addition, we found that one participant tried to change tags in the current tagging application by changing tags in the drop-down menu.

We also observed that most participants(10) linearly watched the video and performed tagging the first time they watched, while four participants used the transcript time code to jump around the video. One participant watched a section of the video and then went back to tag. Eight participants rewatched the video and reviewed their tagging applications, where four of them first used the filmstrip to navigate the tagging application and then browse content on the transcript. Two used the transcript time code to browse tagging content. One participant searched key words in the transcript or used the time code to browse tagging content. P7 did not tag at all but used the filmstrip to review video content.

We allowed participants to use the quick tagging interface as a cheat sheet during the quiz. We observed that seven participants browsed the transcript or used time code to jump around the transcript, and sometimes even referred to their tagging applications. Four participants used the filmstrip to navigate the tagging application and browse content on the transcript. Three participants browsed the transcript or searched using key words.

Other Findings

In our interviews, two participants preferred taking notes while watching educational videos. One suggested that a notepad could be attached to the video interface, where notes themselves could be tagged. One participant liked both highlighting and taking notes to learn with videos. Six participants indicated that they like to highlight on PDF while learning, so they preferred highlighting when watching educational videos. Three of them did mention that colour coding tags were acceptable. However, two participants indicated that they like to tag the video content while learning.

Discussion

From the results, we can see that participants do not agree on using the modes or context menu to add tags and edit their tagging applications. We can say that each design has both advantages and disadvantages. In regards to tagging modes, we suggest to use these tagging modes to add tags with the two different orders, and to use the untagging mode to edit or delete tagging applications. The advantage of using these modes is that it can be integrated with other annotation methods such as highlighting and commenting to create a comprehensive video annotation system. However, the disadvantage is that it may take extra effort for users to change between these two modes, especially when they are new to the video topic as well as the video interface.

In terms of the context menu, users can benefit from performing all tagging actions in one menu, without adding extra cognitive load. But this approach may not be scalable when there are other annotation methods involved, which may cause the menu to become cluttered.

When learners are new to the video topic or when they are under a time-constricted context, we see that they prefer to use transcript tagging. But if the transcript is poor, or if there are graphs or charts in the filmstrip, they prefer using filmstrip tagging so that they can quickly find and recall content. For playhead tagging, it requires very low click effort, but it is not so useful when people need highly accurate information, such as for a quiz or exam situation. However, this may be useful for students to preview video content.

In summary, tagging can useful but is not the only optimal way for students to learn with video annotation. Highlighting and taking notes can also be useful for video annotation to adapt to different learning habits of different learners. For people who strongly prefer highlighting, combining tagging with colour can be a solution. Whereas for people who strongly prefer notes, tagging notes would be beneficial.

5.2.3 Further Improvement to the Tagging Interface

Eight participants mentioned that they would like to save their tag history and filter their tagging content by tags. There are two suggestions from P4 and P5 on this filtering function.

> "A filtering option to go through, looking at only sections of the video tags as "difficult", for example, would be very helpful." [P4]

> "It would be nice to have an "overall" view of the tags I used without having to manually search them in the transcript, playhead, filmstrip, etc.. Maybe use linked highlighting like if I hover over a tag anywhere on the UI, it will highlight wherever else that tag might exist (eg. filmstrip, in transcript, etc.)." [P5]

Two participants suggested that it would be better to have a tag cloud around corresponding video content so that they could have a better understanding of each video section. Three participants mentioned that the drop-down menu slowed down their process of changing tags and blocked their vision for video watching. Therefore, they suggested a horizontal layout. Another two participants liked the idea of customizing the drop-down menu. More specifically, they wanted the menu to simply show their own most-used tags. In regards to the transcript, two participants mentioned that it would be time consuming to browse tagging applications on the transcript for long videos. Another participant commented that the transcript should have larger font sizes. In terms of the filmstrip, one participant suggested that it should be segmented into small sections of real video instead of a number of frames. Finally, one participant would prefer to use the mainviewer to tag in a real scenario, such as using a right click method and choosing tags.

Discussion

As our experiment mainly focuses on adding tags, other functions like filtering and reviewing tagging history can be further improved. The suggestions on these functions from participants are quite useful. For iterative design of the drop-down menu, this could be horizontal and customizable, where users can input and customize their own tags. For topic tags, they could be visualized dynamically in a tag cloud. In addition, reaction tags could be transformed into emojis. We also need to explore further visualization methods for filmstrip itself. Finally, tagging on the mainviewer is yet another future research goal.

5.3 Summary

This chapter introduced a quick tagging mechanism which helped users to recall video content. We performed a controlled study to quantitatively compare time spent using three quick tagging interface with time spent using hand-written notes. We found no significant difference in finishing the quiz between the two methods, but we did find that participants using the quick tagging interface spent around 10% longer on average to watch and learn the video than when using hand-written notes. In terms of reaction to our quick tagging mechanism, users significantly agreed that it was helpful for them to recall content and finish their learning tasks. We also qualitatively collected a substantial number of suggestions on how to improve the quick tagging mechanism, in regards to the pre-defined tags, the quick tagging interaction and modes, the tagging interface visualization, and future designs.

Chapter 6

Conclusion and Future Work

In this chapter, we present a conclusion of the results of our two studies to validate that our proposed quick tagging mechanism can indeed enhance the video learning experience. We also discuss the significance and contributions of this thesis, as well as a proposal for future research.

6.1 Conclusion

In this thesis, we presented an answer to our research questions:

1) Do users feel it's efficient and useful enough to perform quick tagging on video content when finishing their learning tasks? 2) In perspective of usefulness, does quick tagging mechanism help students recall video content?

Overall, participants have a positive reaction to our design. They were able to efficiently bookmark and recall video content of this study. We completed content analysis on course video comments from YouTube and CLAS material, and validated the results in an online survey. We obtained implications for how to choose optimal pre-defined tags under learning contexts, as well as a list of commonly used tags. We then leveraged our quick tagging mechanism, which is comprised of pre-defined tags, video part selection, and applying tags on a video interface. The tagging interface was both qualitatively and quantitatively evaluated in a controlled study. Participants using our quick tagging interface spent around 10% longer on average than when taking notes on paper while learning and watching the video. In terms of quiz completion time, it took participants almost the same average time as taking notes on paper. It can be explained that taking notes engaged them at a different cognitive level to synthesize the content, while the quick tagging helped with other aspects of Bloom's taxonomy using a different type of effort. In addition, there were some tradeoffs involved when using tagging. It took users less time using tagging when they want to quickly mark learning content compared with taking notes. However, notes provide more information than tags when users review their recorded video content. If these two methods are combined together, the combination will likely be useful and beneficial.

In summary, our observations with our tagging interface reinforce the concept that there is not one best way for students to annotate video content while learning. Rather, students learn in different ways at different stages for different content and types of information. Thus, providing a suite of different tools for video annotation, matched to different styles and types of information for learning seems appropriate. Our results suggest that our quick tagging interface for video can be used effectively as one of these tools.

6.2 Future Work

In this thesis, the tagging interface was tested in a controlled lab study and with a limited number of participants. The next step in evaluating this quick tagging mechanism is to expand on this controlled study, and to build a full fledged application so that the tagging function can be used in real learning scenarios. To perform a field study of this magnitude, we plan to integrate this quick tagging feature into our personalized video learning tool, ViDex. This will allow us to more easily test our quick tagging feature with students in their learning contexts.

Although our work is focused on exploring personalized learning with video, it can be a good foundation for building aggregated video learning systems. Besides the traditional interaction data such as play, pause, and seek, tags and tagged intervals can be fed back to the interface design to show the wisdom of the crowd. For example, a tag cloud of specific video parts, or the whole video, can be beneficial for students to pay attention to certain important parts of the course as well as for instructors to better understand student learning feedback. In addition, our current tags are in text format which can be a good foundation for expanding to other more sophisticated content, such as emoji. Other media such as voice can also be explored using our current quick tagging mechanism in the future.

As for our interface features, we included the mainviewer, transcript, and filmstrip together. But for different educational video production styles, the interface features may not adapt well to certain video formats. For example, the filmstrip may not be useful for a video mostly showing an instructor talking. And thus, how to adapt the quick tagging mechanism to other interface features can be further explored. In addition, our current tagging applications are synchronized between transcript and filmstrip. A solution to synchronize them with more interface elements can indeed be a challenge.

The contributions in this work have shown that providing users the quick tagging mechanism is beneficial for effective video learning. Providing optimal pre-defined tags, and proper methods of video part selection and tag application is an essential part of bookmarking video content.

Bibliography

[1] Gregory D. Abowd, Matthias Gauger, and Andreas Lachenmann. The family video archive: An annotation and browsing environment for home movies. In *Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*, MIR '03, pages 1–8, New York, NY, USA, 2003. ACM.

[2] David Bargeron, Jonathan Grudin, Anoop Gupta, and Elizabeth Sanocki. Annotations for streaming video on the web: System design and usage studies. Technical report, March 1999.

[3] Patrick Baudisch. Using a painting metaphor to rate large numbers of objects. In *Proceedings of HCI International (the 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Ergonomics and User Interfaces*, pages 266–270. ACM, August 1999.

[4] CJ Brame. Effective educational videos. *Vanderbilt University Center for Teaching*, 2015.

[5] Bloom B.S. and Krathwohl D.R. *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* 1956.

[6] W. J. Conover. *Practical Nonparametric Statistics, 3rd Edition.* John Wiley & Sons, Inc, 1999.

[7] Nicholas Diakopoulos and Irfan Essa. I.: Videotater: an approach for pen-based digital video segmentation and tagging. In *In UIST 06: Proceedings of the 19th annual ACM symposium on User interface software and technology (New*, pages 221–224. ACM Press.

[8] Nicholas Diakopoulos and Irfan Essa. Videotater: an approach for pen-based digital video segmentation and tagging. In *UIST '06 Proceedings of the 19th annual ACM symposium on User interface software and technology*, pages 221–224. ACM, October 2006.

[9] W E. Hick. On the rate of information gain. 4:11–26, 03 1952.

[10] Matthew Fong, Gregor Miller, Xueqin Zhang, Ido Roll, Christina Hendricks, and Sidney Fels. An investigation of textbook-style highlighting for video. In *Proceedings of Graphics Interface 2016*, GI 2016, pages 201–208. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, 2016.

[11] Scott Golder and Bernardo A. Huberman. The Structure of Collaborative Tagging Systems. *Journal of Information Science*, 32(2):198–208, 2006.

[12] Philip J. Guo, Juho Kim, and Rob Rubin. How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 41–50, New York, NY, USA, 2014. ACM.

[13] Kyungsik Han, Jin Y. Jang, and Dong W. Lee. Exploring tag-based like networks. In *proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA'15, pages 1941–1946. ACM, 2015.

[14] Juho Kim, Elena L. Glassman, Andrés Monroy-Hernández, and Meredith Ringel Morris. Rimes: Embedding interactive multimedia exercises in lecture videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1535–1544, New York, NY, USA, 2015. ACM.

[15] Juho Kim, Philip J. Guo, Carrie J. Cai, Shang-Wen (Daniel) Li, Krzysztof Z. Gajos, and Robert C. Miller. Data-driven interaction techniques for improving navigation of educational videos. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 563–572, New York, NY, USA, 2014. ACM.

[16] Kirsty Kitto, Sebastian Cross, Zak Waters, and Mandy Lupton. Learning analytics beyond the lms: the connected learning analytics toolkit. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, LAK '15, pages 11–15. ACM, 2015.

[17] Anna Korhonen and Ted Briscoe. Extended lexical-semantic classification of english verbs. In *Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics*, CLS '04, pages 38–45. ACM, 2004.

[18] David R. Krathwohl. A revision of bloom's taxonomy: An overview. *Theory Into Practice*, 41(4):212–218, 2002.

[19] George Macgregor and Emma McCulloch. Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, 55(5):291–300, 2006.

[20] W. E. Mackay. Eva: an experimental video annotator for symbolic analysis of video data. In *ACM SIGCHI Bulletin*, pages 68–71. ACM, October 1989.

[21] Toni-Jan Keith Palma Monserrat, Shengdong Zhao, Kevin McGee, and Anshul Vikram Pandey. Notevideo: Facilitating navigation of blackboard-style lecture videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1139–1148, New York, NY, USA, 2013. ACM.

[22] Mitchell J. Morris. *VastMM-Tag: Semantic Indexing and Browsing of Videos for E-Learning*. PhD thesis, Columbia University, New York, United States, 2012.

[23] Xiangming Mu. Towards effective video annotation: An approach to automatically link notes with video content. *Computers & Education*, 55(4):1752 – 1763, 2010.

[24] Katja Niemann. Increasing the accessibility of learning objects by automatic tagging. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, LAK '15, pages 414–415. ACM, 2015.

[25] Amy Pavel, Colorado Reed, Björn Hartmann, and Maneesh Agrawala. Video digests: A browsable, skimmable format for informational lecture videos. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 573–582, New York, NY, USA, 2014. ACM.

[26] Jamie Pina, Kelley Chester, Diana Danoff, and Mark Koyanagi. Synonym-based word frequency analysis to support the development and presentation of a public health quality improvement taxonomy in an online exchange. *Stud Health Technol Inform*, 192:1128–1128, 2013.

[27] Gonzalo Ramos and Ravin Balakrishnan. Fluid interaction techniques for the control and annotation of digital video. In *Proceedings of the 16th*

Annual ACM Symposium on User Interface Software and Technology, UIST '03, pages 105–114, New York, NY, USA, 2003. ACM.

[28] Michael Riegler, Mathias Lux, Vincent Charvillat, Axel Carlier, Raynor Vliegendhart, and Martha Larson. Videojot: A multifunctional video annotation tool. In *Proceedings of International Conference on Multimedia Retrieval*, ICMR '14, pages 534:534–534:537, New York, NY, USA, 2014. ACM.

[29] Evan F Risko, Tom Foulsham, Shane Dawson, and Alan Kingstone. The collaborative lecture annotation system (clas): A new tool for distributed learning. *IEEE Transactions on Learning Technologies*, 6(1):4–13, 2013.

[30] Yvonne Rogers, Helen Sharps, and Jenny Preece. *Interaction Design: Beyond Human - Computer Interaction, 3rd Edition*. Wiley Publishing, 2011.

[31] Schulte im Walde Sabine. Human Associations and the Choice of Features for Semantic Verb Classification. *Res on Lang and Comput*, 6(1):79–111, 2008.

[32] Klaus R. Scherer. What are emotions? And how can they be measured? *Social Science Information*, 44(4):695–729, 2005.

[33] Shilad Sen, Shyong K. Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F. Maxwell Harper, and John Riedl. Tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, CSCW '06, pages 181–190, New York, NY, USA, 2006. ACM.

[34] Margaret A. Storey, Li T. Cheng, Ian Bull, and Peter Rigby. Waypointing and social tagging to support program navigation. In *Ext. Abstracts CHI 2006*, pages 1367–1372. ACM Press, 2006.

[35] Anthony Tang and Sebastian Boring. #epicplay: Crowd-sourcing sports video highlights. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1569–1572, New York, NY, USA, 2012. ACM.

[36] J. Trant. Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. *New Review in Hypermedia and Multimedia*, 12(1):83–105, 2006.

[37] Lex V. Velsen and Mark Melenhorst. Incorporating user motivations to design for video tagging. *Elsevier B.V.*, 21(3):221–232, 2009.

[38] Timo Volkmer, John R. Smith, and Apostol(Paul) Natsev. A web-based system for collaborative annotation of large image and video collections: an evaluation and user study. In *MULTIMEDIA '05 Proceedings of the 13th annual ACM international conference on Multimedia*, pages 892–901. ACM, November 2005.

[39] Alex Y. Zheng, Janessa K. Lawhorn, Thomas Lumley, and Scott Freeman. Application of bloom's taxonomy debunks the "mcat myth". *Science*, 319(5862):414–415, 2008.

[40] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. Successful classroom deployment of a social document annotation system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1883–1892. ACM, 2012.

Appendix A

Online Survey Questionnaire

ViDex Research

Page #1

Survey on Educational Video Tagging Interface by the Human Communication Technologies Laboratory
University of British Columbia
UBC Ethics Certificate Number: H13-01589
Contact: Xueqin Zhang, xueqin@ece.ubc.ca Thank you for taking part in our ViDeX Research survey on video tagging. The purpose of this survey is to collect words for a video tagging interface. The data will be used to inform the design of a video tagging tool to allow people to manage, visualize, and utilize their video course notes in a form of tags for video parts. You will be asked to watch
eo clips as well as to rate 3 default tag layouts. The whole survey will take you about 10 mins. We need
ead and sign the Consent Form. Fill in your demographic information. Describe the video clips with
eedback on default tag layouts. Answer some questions on the diffiulcty of the tagging task.

Page #2

THE UNIVERSITY OF BRITISH COLUMBIA
Department of Electrical Computer Engineering
2332 Main Mall
Vancouver, B.C., V6T 1Z4

August 30, 2016
Consent Form Principal Investigator

Dr. Sidney Fels, Professor, Department of Electrical and Computer Engineering, University of British Columbia (604) 822-5338

Co-Investigators

Xueqin Zhang, Matthew Fong, Gregor Miller

Project Purpose and Procedures

This project is designed to understand how users take notes through a video space and to investigate how a video part tagging interface could help users navigate and summarize this video. The purpose of this study is to discover and aid users in studying for courses that use video resources. You will be asked to watch and describe 5 course video clips as well as to rate 3 default tag layouts. The data will be used to inform the design of a video part tagging tool to allow people to manage, visualize, and utilize their video course notes in a form of tags for video parts. You may stop your participation at any time.

Confidentiality

The data gathered will be anonymised for the purposes of subsequent analysis, however, you will be in contact with others during the study. Identifiable data will be stored securely in a password protected computer account. All data from individual participants will be coded so that their anonymity will be protected in any project reports and presentations that result from this work.

Contact Information About the Project

If you have any questions or require further information about the project you may contact Dr. Sidney Fels, 604-822-5338

Contact for information about the rights of research subjects

If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the UBC Office of Research Ethics at 604-822-8598 or if long distance email RSIL@ors.ubc.ca or call toll free 1-877-822-8598. The ethics approval ID for this project is H13-01589.

Consent

We intend for your participation in this project to be pleasant and stress-free. Your participation is entirely voluntary and you may refuse to participate or withdraw from the study at any time.
Your agreement indicates that you consent to participate in this project. You do not waive any legal rights by agreeing this consent form.

- I agree to participate in the project as outlined above. My participation in this project is voluntary and I understand that I may withdraw at any time.

______/____/____(YYYY/MM/DD)

Page #3

Gender

- Female
- Male

Age

- less than 19
- 19 - 25
- 26 - 30
- 31 - 40
- 41 - 50
- 51 - 60
- 61 or over

Are you an English native speaker?

- Yes
- No

Please specify your Faculties and Schools

- Applied Science
- Business, Sauder School
- Architecture and Landscape Architecture School
- Arts
- Audiology and Speech Sciences School
- Community and Regional Planning School
- Continuing Studies
- Dentistry
- Education
- Forestry

... 3 additional choices hidden ...

- Law, Peter A. Allard School
- Library, Archival and Information Studies School
- Medicine

- Music School
- Nursing School
- Population and Public Health School
- Pharmaceutical Sciences
- Science
- Social Work School
- UBC Vantage College

How many courses have you taken at UBC that taught with video?

- 1 - 2
- 3 - 5
- 5+
- None

How many online courses have you taken that taught with video?

- 1 - 2
- 3 - 5
- 5+
- None

Which online services have you taken courses from?

- Cousera
- edX
- Khan Academy
- MIT OpenCourseWare
- Udacity
- Other (specify) ________________________

Page #4

In this section, there are 5 twenty-second length course video clips in total. They cover five subjects: music, art history, library, math and engineering. Firstly, you need to watch the video clip and then tag it. We will provide you with some words to choose from. But you are encouraged to input your own words.

When you are ready, please click Next to start the test.

Page #5

Video 1

Page #6

(Multiple Choices) Do you think which words in the following can best express your feelings about watching Video1?

- good

- like
- easy
- None of the given words apply
- enjoy
- bad
- amazing
- awesome
- comfortable
- interesting

(Optional) Is the word you want not there? Put them in here! (for more than one word, separate each with a comma)

(Multiple Choices) Do you think which words in the following can properly describe the content of Video1?

- true
- clear
- helpful
- None of the given words apply
- new
- important
- wrong
- precise
- limited
- pretty

(Optional) Is the word you want not there? Put them in here! (for more than one word, separate each with a comma)

Page #7

Video 2

Page #8

(Multiple Choices) Do you think which words in the following can best express your feelings about watching Video2?

- good
- like
- easy
- None of the given words apply
- enjoy
- bad
- amazing
- awesome
- comfortable
- interesting

(Optional) Is the word you want not there? Put them in here! (for more than one word, separate each with a comma)

(Multiple Choices) Do you think which words in the following can properly describe the content of Video2?

- true
- clear
- helpful
- None of the given words apply
- new
- important
- wrong
- precise
- limited
- pretty

(Optional) Is the word you want not there? Put them in here! (for more than one word, separate each with a comma)

Page #9

Video 3

Page #10

(Multiple Choices) Do you think which words in the following can best express your feelings about watching Video3?

- good
- like
- easy
- None of the given words apply
- enjoy
- bad
- amazing
- awesome
- comfortable
- interesting

(Optional) Is the word you want not there? Put them in here! (for more than one word, separate each with a comma)

(Multiple Choices) Do you think which words in the following can properly describe the content of Video3?

- true
- clear
- helpful
- None of the given words apply
- new
- important
- wrong
- precise
- limited
- pretty

(Optional) Is the word you want not there? Put them in here! (for more than one word, separate each with a comma)

Page #11

Video 4

Page #12

(Multiple Choices) Do you think which words in the following can best express your feelings about watching Video4?

- good
- like
- easy
- None of the given words apply
- enjoy
- bad
- amazing
- awesome
- comfortable
- interesting

(Optional) Is the word you want not there? Put them in here! (for more than one word, separate each with a comma)

____

(Multiple Choices) Do you think which words in the following can properly describe the content of Video4?

- true
- clear
- helpful
- None of the given words apply
- new
- important
- wrong
- precise
- limited
- pretty

(Optional) Is the word you want not there? Put them in here! (for more than one word, separate each with a comma)

____

Page #13

Video 5

Page #14

(Multiple Choices) Do you think which words in the following can best express your feelings about watching Video5?

- good
- like
- easy
- None of the given words apply
- enjoy
- bad
- amazing
- awesome
- comfortable
- interesting

(Optional) Is the word you want not there? Put them in here! (for more than one word, separate each with a comma)

____

(Multiple Choices) Do you think which words in the following can properly describe the content of Video5?

- true
- clear
- helpful
- None of the given words apply
- new
- important
- wrong
- precise
- limited
- pretty

(Optional) Is the word you want not there? Put them in here! (for more than one word, separate each with a comma)

____

Page #15

Now, we have three layouts for displaying tags. The layouts will be shown as drawings. You will be asked to rate the layouts on a seven-point scale and to anwser some questions.

When you are ready, please click Next to begin.

Page #16

How do you like the layout above?

	1(dislike at all)	2	3	4(neutral)	5	6	7(like very much)
The phrases used to categorize the words	○	○	○	○	○	○	○
The click interaction to show the words	○	○	○	○	○	○	○

Organize the words into three categories	○	○	○	○	○	○	○

Do you think what is the most proper number to display the words in each category?

- ○ 1
- ○ 3
- ○ 5
- ○ 7
- ○ Other, please specify... ________________________

Page #17

How do you like the layout above?

	1(dislike at all)	2	3	4(neutral)	5	6	7(like very much)
The phrases used to categorize the words	○	○	○	○	○	○	○
The click interaction to show the words	○	○	○	○	○	○	○
Organize the words into three categories	○	○	○	○	○	○	○

Do you think what is the most proper number to display the words in each category?

- ○ 1
- ○ 3
- ○ 5
- ○ 7
- ○ Other, please specify... ________________________

Page #18

How do you like the layout above?

	1(dislike at all)	2	3	4(neutral)	5	6	7(like very much)
The way to show all the words	○	○	○	○	○	○	○

If you have any other suggestions about our default tag layout design, please write down below.

____
____
____

Page #19

Please score the amount of effort it took you to add tags for the video clips.

- very much
- a lot
- reasonable amount
- a little
- not at all

Are the words we provide useful to describe video content?

- not at all
- a little
- reasonable amount
- a lot
- very much

Are the words we provide useful to express your feelings about watching videos?

- not at all
- a little
- reasonable amount
- a lot
- very much

Which way do you like most to describe video content?

- use default words
- input your own words
- don't care about
- Other, please specify... ________________________

Which way do you like most to express your feelings about watching videos?

- use default words
- input your own words
- don't care about
- Other, please specify... ________________________

Page #20

Thanks for completing the survey.

Appendix B

Lab Study Questionnaire

Page #1

THE UNIVERSITY OF BRITISH COLUMBIA
Department of Electrical Computer Engineering
2332 Main Mall
Vancouver, B.C., V6T 1Z4

August 7, 2017 Consent Form Principal Investigator Dr. Sidney Fels, Associate Professor, Department of Electrical and Computer Engineering, University of British Columbia (604) 822-5338
Co-Investigators Xueqin Zhang, Matthew Fong
Project Purpose and Procedures This project is designed to understand how users learn and annotate through a video space and to investigate how a quick tagging interface could help users navigate this video and recall video content. The purpose of this study is to discover and aid users in studying for courses that use video resources. You will be given a course video to watch and learn. The video is segmented into two parts. You will be asked to learn one part with no tagging inferface and finish the corresponding quiz. You will be also asked to finish a proficiency test for our quick tagging interface in order to proceed to the other part with tagging interface. You are encouraged to use our quick tagging interface to mark down the video content to finish a quiz afterwards. After finishing the quiz, we will ask about your use experience by a questionnaire. Finally, you will be asked to freely explore our quick tagging interface and we will ask about your use experience by a follow-up interview. The data will be used to inform the design of a video tagging tool to allow people to manage, visualize, and utilize their video course content in a form of tags for video parts. You may stop your participation at any time.
Confidentiality The study involves the audio recording of your interview with the researcher and the video capture of the computer screen to record your use patterns. Neither your name or any other identifying information will be associated with the audio or audio recording or transcript. Only the research team will be able to listen (view) to the recordings. The data gathered will be anonymised for the purposes of subsequent analysis, however, you will be in contact with others during the study. Identifiable data will be stored securely in a password protected computer account. All data from individual participants will be coded so that their anonymity will be protected in any project reports and presentations that result from this work.
Contact Information About the Project If you have any questions or require further information about the project you may contact Dr. Sidney Fels, 604-822-5338
Contact for information about the rights of research subjects If you have any concerns or complaints about your rights as a research participant and/or your experiences while participating in this study, contact the Research Participant Complaint Line in the UBC Office of Research Ethics at 604-822-8598 or if long distance email RSIL@ors.ubc.ca or call toll free 1-877-822-8598. The ethics approval ID for this project is H13-01589.
Consent We intend for your participation in this project to be pleasant and stress-free. Your participation is entirely voluntary and you may refuse to participate or withdraw from the study at any time.
Your agreement indicates that you consent to participate in this project. You do not waive any legal rights by agreeing with this consent form.

- I am allowing the researcher to audio me or video capture my screen as part of this research.

- I agree to participate in the project as outlined above. My participation in this project is voluntary and I understand that I may withdraw at any time.

______/____/____(YYYY/MM/DD)

Page #2

Please choose the most appropriate selection:

Gender

- Female
- Male

Age

- less than 19
- 19 - 25
- 26 - 30
- 31 - 40
- 41 - 50
- 51 - 60
- 61 or over

Are you an English native speaker?

- Yes
- No

Please specify your Faculties and Schools

• Applied Science
• Business, Sauder School
• Architecture and Landscape Architecture School
• Arts
• Audiology and Speech Sciences School
• Community and Regional Planning School
• Continuing Studies
• Dentistry
• Education
• Forestry
... 3 additional choices hidden ...
• Law, Peter A. Allard School
• Library, Archival and Information Studies School
• Medicine
• Music School
• Nursing School
• Population and Public Health School
• Pharmaceutical Sciences
• Science
• Social Work School
• UBC Vantage College

How many courses have you taken at UBC that taught with video?

- 1 - 2
- 3 - 5
- 5+
- None

How many online courses have you taken that taught with video?

- 1 - 2
- 3 - 5
- 5+
- None

Which online services have you taken courses from?

You can choose more than one selection if applicable.

- Cousera
- edX
- Khan Academy
- MIT OpenCourseWare
- Udacity
- Other (specify) ______________________

Page #3

In this section, there are some questions to ask you about your use experience on the quick tagging interface.

When you are ready, please click Next to start the questionnaire.

Page #4

Please rate agreement or disagreement with the following statements:
Where 1 represents strongly disagree and 5 represents strongly agree

The quick tagging interface is

	1(strongly disagree)	2(disagree)	3(neutral)	4(agree)	5(strongly agree)
Easy to use	○	○	○	○	○
Efficient to tag video content	○	○	○	○	○
Useful to recall video content	○	○	○	○	○

The words provided in the drop-down menu are

	1(strongly disagree)	2(disagree)	3(neutral)	4(agree)	5(strongly agree)
Helpful to recall video content	○	○	○	○	○
Useful to mark down video content	○	○	○	○	○

I like the way to tag a video part

	1(strongly disagree)	2(disagree)	3(neutral)	4(agree)	5(strongly agree)
in the transcript	○	○	○	○	○
in the filmstrip	○	○	○	○	○
in the playhead	○	○	○	○	○

The quick tagging mechanism is helpful to finish my learning tasks while watching the course video

- ○ 1(strongly disagree)
- ○ 2(disagree)
- ○ 3(neutral)

- 4(agree)
- 5(strongly agree)

Please add any recommendations for changes to the overall design, tags or navigation of the tagging interface. Your comments are very important for us.

Thanks for your participation in the testing of this prototype.

____
____
____

Please score the amount of effort it took you to tag the video content.

	1(not at all)	2(somewhat)	3(reasonable amount)	4(a lot)	5(very much)
Tagging in the transcript	○	○	○	○	○
Tagging in the filmstrip	○	○	○	○	○
Tagging in the playhead	○	○	○	○	○

Please score your familarity with the video topic.

- 1(not at all)
- 2(somewhat)
- 3(reasonable amount)
- 4(a lot)
- 5(very much)

Please score your interest with the video topic.

- 1(not at all)
- 2(somewhat)
- 3(reasonable amount)
- 4(a lot)
- 5(very much)

Appendix C

Intermediate Results of Background Study

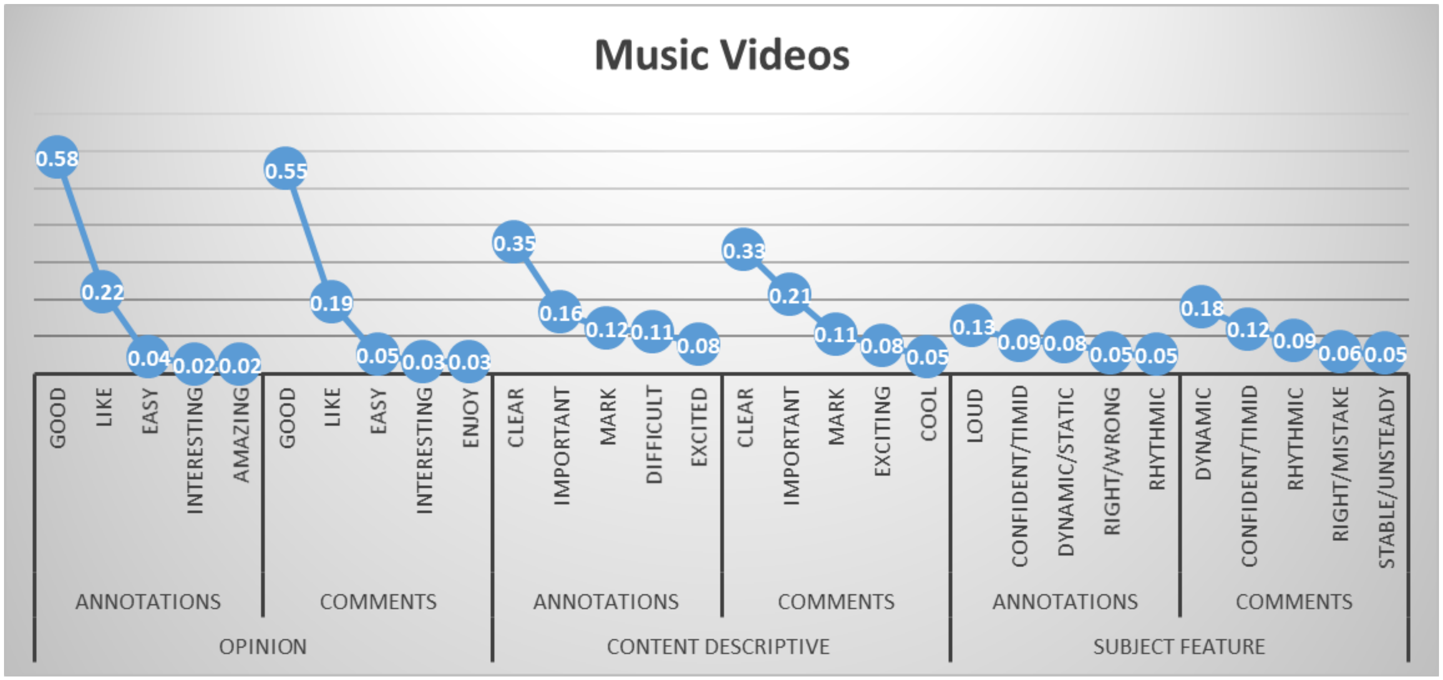

Figure C.1: Comparison results between annotations and comments in music courses from CLAS. The blue bubbles show normalized percentages. The comparison was run among three categories: opinion words, content descriptive words and subject feature words.

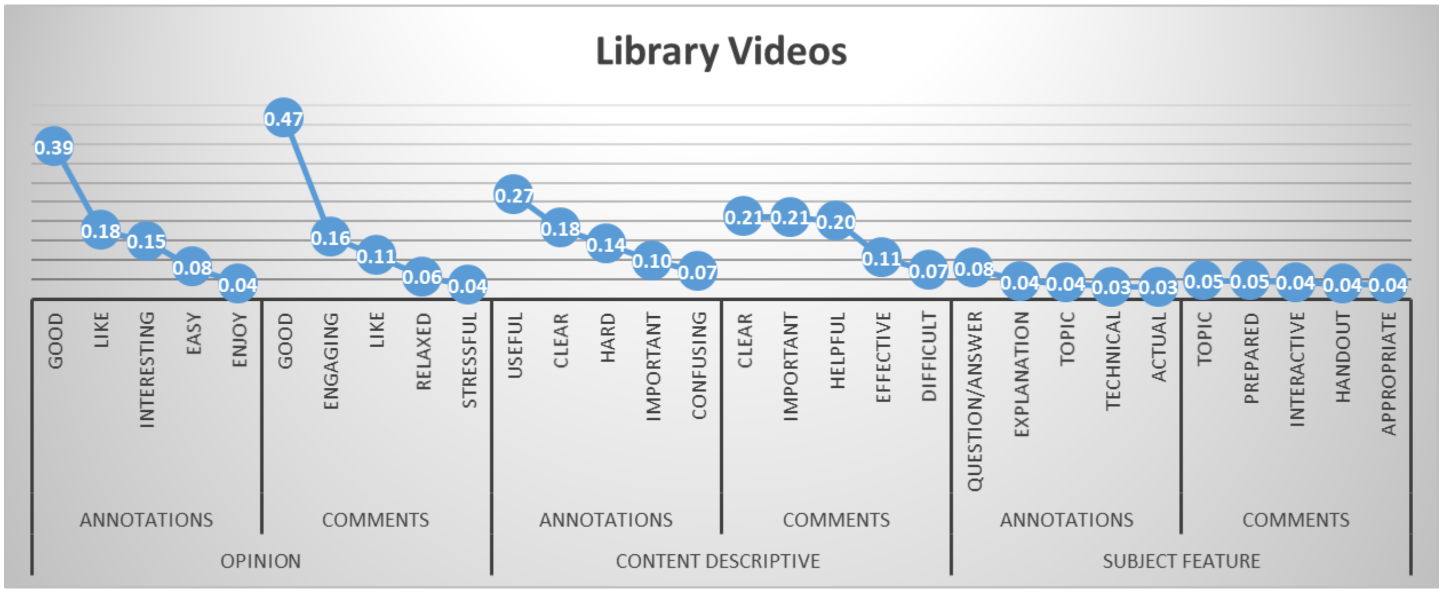

Figure C.2: Comparison results between annotations and comments in library courses from CLAS. The blue bubbles show normalized percentages. The comparison was run among three categories: opinion words, content descriptive words and subject feature words.

Appendix D

Aggregation Process of Background Study

Opinion Words

CLAS			Youtube		
word	count	percentage	word	count	percentage
good	486	0.465517241	good	1234	0.330299786
like	177	0.16954023	like	1020	0.273019272
easy	128	0.122605364	interesting	234	0.062633833
interesting	88	0.084291188	happy	207	0.055406852
enjoy	20	0.019157088	bad	203	0.054336188
comfortable	18	0.017241379	amazing	172	0.046038544
nervous	18	0.017241379	awesome	141	0.037740899
awesome	15	0.014367816	simple	72	0.019271949
bad	13	0.012452107	funny	60	0.016059957
intimidated	13	0.012452107	hate	51	0.013650964
funny	12	0.011494253	fantastic	41	0.010974304
amazing	11	0.010536398	depressed	34	0.009100642
worried	10	0.009578544	satisfied	31	0.008297645
pleased	10	0.009578544	horrible	31	0.008297645
moved	8	0.007662835	boring	31	0.008297645
fantastic	7	0.006704981	dislike	27	0.007226981
boring	2	0.001915709	sad	27	0.007226981
awkward	2	0.001915709	disappointed	26	0.006959315
hopeful	1	0.000957854	moving	17	0.004550321
doomed	1	0.000957854	joyful	17	0.004550321
revolt	1	0.000957854	pleasant	17	0.004550321
disappointed	1	0.000957854	hopeful	9	0.002408994
freaking	1	0.000957854	uncomfortable	9	0.002408994
embarrassment	1	0.000957854	angry	7	0.001873662
			awkward	5	0.00133833
			annoying	5	0.00133833

			embarrassing	5	0.00133833
			worrying	3	0.000802998
		good	0.465517241		
		like	0.273019272		
		easy	0.122605364		
		interesting	0.084291188		
		enjoy	0.055406852		
		bad	0.054336188		
		amazing	0.046038544		
		awesome	0.037740899		
		comfortable	0.017241379		
Content Descriptive Words					
clear	126	0.12962963	true_	583	0.24352548
helpful	110	0.113168724	important	145	0.060568087
right	97	0.099794239	wrong	144	0.060150376
new	86	0.088477366	pretty	125	0.052213868
important	68	0.069958848	clear	103	0.043024227
precise	56	0.057613169	helpful	102	0.042606516
limited	51	0.052469136	hard	94	0.039264829
difficult	37	0.038065844	original	83	0.034670008
pretty	37	0.038065844	intelligent	82	0.034252297
agree	35	0.03600823	favorite	80	0.033416876
appropriate	28	0.028806584	nonsense	69	0.028822055
mark	27	0.027777778	deep	64	0.0267335
primary	27	0.027777778	believe	58	0.024227235
understandable	18	0.018518519	cool	54	0.022556391
mistake	17	0.017489712	reasonable	48	0.020050125
confusing	16	0.016460905	proper	48	0.020050125
effective	16	0.016460905	accurate	46	0.019214703
excited	15	0.015432099	crazy	44	0.018379282
complex	15	0.015432099	inspiring	42	0.01754386
lovely	11	0.011316872	oppose	41	0.017126149
comprehensive	10	0.010288066	confusing	36	0.015037594
decent	10	0.010288066	appreciate	35	0.014619883
attentive	8	0.008230453	evil	35	0.014619883
brief	7	0.007201646	skeptical	31	0.012949039
benefited	7	0.007201646	complex	29	0.012113617
wise	7	0.007201646	inaccurate	25	0.010442774
tricky	5	0.005144033	limited	24	0.010025063
crazy	5	0.005144033	support	23	0.009607352

primitive	5	0.005144033	insightful	19	0.007936508
convincing	4	0.004115226	excited	17	0.007101086
common	2	0.002057613	impressive	11	0.00459482
concrete	2	0.002057613	effective	11	0.00459482
careful	2	0.002057613	understandable	8	0.003341688
expressive	1	0.001028807	concrete	5	0.002088555
unsure	1	0.001028807	special	5	0.002088555
unreality	1	0.001028807	ugly	5	0.002088555
unproblematic	1	0.001028807	tricky	4	0.001670844
unexpressive	1	0.001028807	exhausting	4	0.001670844
unproblematic	1	0.001028807	unclear	2	0.000835422
unexpressive	1	0.001028807	discouraged	2	0.000835422
			shortsighted	2	0.000835422
			unintelligent	1	0.000417711
			disbelieve	1	0.000417711
			insignificant	1	0.000417711
			unreasonable	1	0.000417711
			unappreciated	1	0.000417711
			unimpressive	1	0.000417711
		true_	0.24352548		
		clear	0.12962963		
		helpful	0.113168724		
		new	0.088477366		
		important	0.069958848		
		wrong	0.060150376		
		precise	0.057613169		
		limited	0.052469136		
		pretty	0.052213868		