Exploring user-adaptive visualization techniques with Multi-Modal Documents

by

Dereck Toker

B.A., University of British Columbia, 2010
M.Sc., University of British Columbia, 2013

Doctoral Thesis Proposal

DEPARTMENT OF COMPUTER SCIENCE

The University of British Columbia

(Vancouver)

Supervisory Committee:

Cristina Conati
Giuseppe Carenini
Tamara Munzner
James Enns

September 2016

© Dereck Toker, 2016
Table of Contents

TABLE OF CONTENTS .................................................................................................................. 2
LIST OF TABLES .......................................................................................................................... 3
LIST OF FIGURES ......................................................................................................................... 4

1. INTRODUCTION ................................................................................................................... 5
   1.1 BACKGROUND AND MOTIVATION .................................................................................. 5
   1.2 GOALS AND OBJECTIVES ......................................................................................... 8
   1.3 PROJECT MILESTONES & OVERVIEW ........................................................................... 9
   1.4 CONTRIBUTIONS ........................................................................................................ 15

2. RELATED WORK .................................................................................................................. 17
   2.1 INDIVIDUAL DIFFERENCES IN INFORMATION VISUALIZATION ........................................ 17
   2.2 USER-ADAPTIVE VISUALIZATION SYSTEMS .......................................................... 21
   2.3 PREDICTING USER CHARACTERISTICS FROM EYE TRACKING DATA .......................... 22

3. MILESTONE 1A: IMPLEMENT EXPERIMENTAL PLATFORM .............................................. 24
   3.1 ACQUIRE MMD DATASET ............................................................................................. 24

4. MILESTONE 1B: DESIGN AND CONDUCT USER STUDY .................................................. 28
   4.1 SELECT APPROPRIATE SET OF USER CHARACTERISTICS .......................................... 28
   4.2 SELECT SUITABLE INTERVENTION PROPERTIES ..................................................... 29
   4.3 DETERMINE STUDY TASKS & DEPENDENT MEASURES ............................................ 31
   4.4 DEFINE REFERENCE COMPLEXITY ........................................................................... 32
   4.5 DATA ANALYSIS ....................................................................................................... 40

5. MILESTONE 2: PREDICTING USER CHARACTERISTICS IN REAL-TIME ............................. 41
   5.1 PROCESS EYE TRACKING DATA .................................................................................. 41

APPENDIX A. TIMELINE AND PUBLICATIONS ........................................................................ 43
   A.1 EXISTING PUBLICATIONS: ......................................................................................... 44
   A.2 PROPOSED PUBLICATIONS ....................................................................................... 45

BIBLIOGRAPHY ......................................................................................................................... 47
List of Tables

**Table 1.** List of cognitive abilities and personality traits that have been previously investigated in relation to visualization effectiveness. * indicates that no significant results were identified for that given user characteristic. .............................................................................................................................................18

**Table 2.** A list of properties we identified that capture different aspects of reference complexity........................................34

**Table 3.** Statistics on the visualization types present in Kong dataset ........................................................................................35

**Table 4.** Statistics on the levels of the primary and secondary keys for the 40 visualizations in the Kong dataset ..........36

**Table 5.** Statistics on the search and query types, as well as the number of targets for the 182 references in the Kong dataset. .............................................................................................................................................39

**Table 6.** Example of gaze features generated from EMDAT: A) Gaze features describing visual processing according to the whole screen; B) Gaze features describing activity within one or more Areas of Interest (AOIs) .................42
List of Figures

Figure 1. Multi-Modal Documents (MMDs) each depicting the presence of a body of text and a visualization. Document sources: (a) Pew Research; (b) European Banking Authority; and (c) The Functional Art. ..............................6

Figure 2. Two references in an MMD each consisting of text and corresponding data points within the visualization. [Source: The Economist - Dec 22, 2012] ..........................................................................................................................7

Figure 3. High level overview of Milestone 1 consisting of two major stages: A. Implementing the Experimental Platform, and B. Designing and Conducting the User Study. Green boxes indicate tasks to be addressed. ..................9

Figure 4. High level overview of Milestone 2. Green boxes indicate tasks or challenges to be addressed. ..............................13

Figure 5. Examples of various visualizations users can select in the system presented in Gotz & Wen [2009] .................21

Figure 6. An example from Gravemeyer [2006] showing: Left - How a set of visualization options are presented to users who must select one for solving a database query; Right - Example of the adaptive version of the system, where some options have been removed and one recommended option is highlighted in red. .................................22

Figure 7. Tasks for implementing the experimental platform. ..........................................................................................24

Figure 8. Example MMD in the Kong dataset, displayed using their interface. Here, one reference has been selected in the text, and the nine corresponding bars in the visualization have been highlighted accordingly. ..............................25

Figure 9. Elements from a source document (left) are selected to generate a single-paragraph-single-viz MMD (right) in the Kong dataset. ........................................................................................................26

Figure 10. Example of fragmentation in a Kong MMD, since it is unclear what is meant by `it matters` (underlined in red). ..........................................................................................................................................27

Figure 11. Overview of the user study indicating tasks that will have to be addressed. .....................................................28

Figure 12. Types of highlighting interventions evaluated in Carenini et al. [2014] ...............................................................30

Figure 13. Two references from the Kong dataset. Highlighting indicates which bars are being referenced respectively. 32

Figure 14. The six visualization types contained in the Kong dataset. ................................................................................34

Figure 15. Data table consisting of 2 columns, visualized using a single bar chart. There are 6 levels in the primary key. 35

Figure 16. Data table consisting of 3 columns. There are 3 levels for the primary key, 2 levels for the secondary key, yielding a total of 6 values. The data is visualized using a grouped bar chart. .........................................................36

Figure 17. Example of a single target reference, described by a locate and identify task. ....................................................38

Figure 18. Example of a reference with 3 targets, described by a browse task and compare task. .................................39

Figure 19. Tasks for conducting machine learning experiments in Milestone 2. .............................................................41

Figure 20. Possible starting set of AOIs (Areas of Interest) that we will define over the MMD. ........................................42
1. Introduction

1.1 Background and Motivation

In the current era of the social-web, with the expansion of the internet of things, and the continuous miniaturization of sensors, data is now being collected at an unbridled rate in many aspects of human life. The data being collected includes information from smart phones, autonomous cars, online games, shopping, web browsing/streaming, personal-health, banking, meteorology, and education, just to name a few. In this world of ever-expanding collection of digital information, information visualizations are a powerful tool that can be employed to support tasks such as the exploration, presentation, discovery, and overall better understanding of the underlying trends lurking in all of this data.

Information visualizations (InfoVis for short) are “visual representations of datasets designed to help people carry out tasks more effectively” [Munzner 2014]. There are many different types of InfoVis (e.g., bar graph, scatter plot, heatmap, etc.) which can range in their degree of complexity and level of interactivity. Despite ongoing efforts to innovate and improve the effectiveness of visualizations, they are typically designed and evaluated following a one size-fits-all model, meaning that they do not take into account the individual differences of the user. However, there is mounting evidence that users’ differences such as cognitive abilities, personality traits, learning abilities, and preferences can significantly influence user performance and satisfaction during information visualization tasks (e.g., [Conati & Maclaren 2008],[Green & Fisher 2010],[Ziemkiewicz et al. 2011],[Toker et al. 2012]). These findings have thus prompted researchers to investigate user-adaptive information visualizations, i.e., visualizations that aim to recognize and adapt to each user’s specific needs. Formally, user-adaptive interaction is the practice of changing the layout or elements of an interface based on the current needs of the user [Schneider-Hufschmidt et al. 1993]. To offer a concrete example, Gotz & Wen designed an interface that recommends a more suitable visualization to users. Recommendations were provided according to patterns detected in user behavior as they interacted in real-time with an initial visualization [Gotz & Wen 2009]. Most existing research on examining the relationship between individual differences and information visualization has been limited to tasks where users interact only with visualizations. The aim of our research is to expand the scope of user-adaptive visualization research to include scenarios where users interact with visualizations accompanied by a body of text that discusses various aspects of the visualization. Namely, our goal is to provide user-adaptive support to processing visualizations embedded in Multi-Modal Documents.

Multi-Modal Documents (or MMDs for short), are documents consisting of a body of text as well as non-textual elements, such as visualizations, animated media, photographs, sketches, or diagrams. MMDs appear in many different forms, e.g., newspapers, blogs, magazines, brochures, slide-shows, text-books, reports, etc., and MMDs present information from many different areas, e.g., current events, education, government, business, research, etc. For the purposes of our research, we limit the scope of MMDs to those whose non-textual elements consist only of static information visualizations. Figure 1 provides an example of MMDs containing static information visualizations, and in the rest of this proposal we use the term MMD to indicate this specific subset of MMD types.
Research has shown that text and graphical modalities are well suited information channels to combine [ETSI 2008]. Moreover, the multimedia principle states that “users learn more deeply from words and pictures than from words alone.” [Mayer 2009]. However, an established problem arising from combined modalities is the split-attention effect. Split-attention occurs when users are required to split their attention between two information sources (e.g., text and visualization), which increases cognitive load and can negatively impact learning [Ayres & Cierniak 2012]. This problem is exacerbated in MMDs where often there is more than one visual task specified through the MMD text. Multiple visual tasks in MMDs are captured by references in the MMD. We define references as pairs where: the first component is a piece of text in the MMD that specifies a visual task in one of the MMD’s visualizations; the second component is the set of items in that visualization that are involved in the visual task. Figure 2 provides an example of two references in an MMD. One involves the sentence “India and China will have further strong rises” and the bars marked by the solid red arrows in the accompanying bar chart. The second reference involves the sentence “Brazil and Britain will suffer reverses” as well as the bars pointed to by the dashed green arrows. Typically, references are used to support arguments or statements being made in the document’s text by providing added details or specific interpretations to an accompanying visualization. As a user reads through an MMD, they will often encounter a variety of references in the text, each soliciting attention to different aspects of the accompanying visualization. However, since the visualizations cannot be designed to favor the performance of any particular reference (because favoring one task may hinder the others), it has been proposed by Carenini et al. [2013] to address this problem using a user-adaptive approach by dynamically highlighting relevant aspects of the visualization depending on what part of the text the user is reading.
The primary focus of our research will consist of expanding on the idea proposed by Carenini et al. [2013] and investigating how to provide effective user-adaptive support for MMD processing. For the purposes of this thesis, the adaptive support will be in the form of interventions that help users process the relevant components of the visualizations embedded in the MMD when they are referred to in the text. This focus will require pursuing the following two objectives:

- Build a **user model** that will track relevant properties of the user necessary for ascertaining when and where to provide adaptive support in processing the visualization embedded in the MMD.
- Identify forms of adaptive **highlighting interventions** on the MMD visualizations that are effective and nonintrusive for the user.

We will investigate the value of adaptations that are triggered when a user looks at reference’s text in the MMD. In order to identify when a reference’s text is looked at, the user model will employ eye tracking to monitor in real-time where the user is looking as they interact with an MMD. Each reference’s text in the MMD refers to a subset of data within the visualization (see Figure 2) thus informing **where** in the visualization the adaptation should be provided. We will investigate forms of adaptive support that rely on **highlighting interventions**, intended to help users by highlighting relevant data within the visualization. Properties of the interventions will include different types of highlighting, as well as how they appear on the screen and fade off the screen. The choice of this type of interventions is informed by previous work by Carenini et al. [2014]. This work showed that drawing/overlaying interventions (e.g., arrows, bolding) which acted as ‘visual prompts’ or ‘cues’ to solicit attention to elements of interest on bar-graph visualizations were successful at improving
visualization effectiveness for users performing sets of standard visualization tasks. We will investigate whether these results generalize to users leveraging visualizations in the context of processing MMDs.

### 1.2 Goals and Objectives

In this PhD research we aim to investigate if, when, and how adaptive support can help users process visualizations as they are referenced in an MMD. We expect that MMDs with user-adaptive interventions will be more effective compared to non-adaptive MMDs, and testing this hypothesis is the main goal of this thesis. More formally, this goal can be expressed as our first research question:

- **RQ1**: To what extent does providing adaptive highlighting interventions, triggered when the user looks at a reference’s text in an MMD, improve the effectiveness of the MMDs?

  Here, effectiveness refers to any number of quantitative and qualitative measures that can be used to evaluate the user’s experience when processing the MMD (e.g., time on task, task accuracy, comprehension, preference, satisfaction).

A mounting body of work (e.g., [Velez et al. 2005],[Conati & Mclaren 2008],[Green & Fisher 2010], [Ottley et al. 2015]) has already shown that individual user differences can play a role in the effectiveness of using a visualization (see Related Work Section 2.1 for more details). Our hypothesis is that the same will be true for interactions with MMDs, which contain visualization(s) along with a body of text that users must process in order to connect the reference text to the relevant items in the visualization they refer to. Therefore, our second research question is:

- **RQ2**: To what extent does the effectiveness of the highlighting interventions in RQ1 depend on the individual differences of the user?

Answering RQ2 is important because it will allow us to extend the user model to include user differences that were shown to impact the effectiveness of different highlighting interventions. By extending the user model in this way, user-adaptive MMDs could be further personalized. I.e., when a user looks at a reference’s text, an optimal highlighting intervention would be selected based on the relevant user differences. Achieving this goal requires that the user’s individual differences can be captured by the user model. An existing body of work has shown that it is possible to ascertain user differences in real-time during the interaction with information visualizations (e.g., [Toker et al. 2014],[Steichen et al. 2014], [Gingerich & Conati 2015],[Lallé et al. 2015]). This existing work involved predicting user differences from eye tracking data while users were performing simple visualization tasks on standalone visualizations (see related work Section 2.3). For this PhD research, we will ascertain if we can find similar results for predicting relevant user differences from eye tracking while users process visualizations embedded in MMDs. Thus, our third research question is:

- **RQ3**: Using eye tracking, to what extent can we reliably predict relevant user characteristics in real-time as the user interacts with MMDs.
Answering RQ3 is important because it aims to provide a non-intrusive way (i.e., with eye tracking) to ascertain relevant user characteristics without having to intrude or occupy the user’s time by administering tests.

1.3 Project Milestones & Overview

Work for this proposal will be presented according to two Milestones. Milestone 1 consists of: A) the work required to develop an experimental platform for rapid prototyping and evaluation of adaptive MMDs; and B) the work required to design and conduct a user study for evaluating our adaptive interventions. Next, Milestone 2 presents work on running machine learning experiments for predicting user differences from eye tracking data that was collected during Milestone 1.

![Figure 3. High level overview of Milestone 1 consisting of two major stages: A. Implementing the Experimental Platform, and B. Designing and Conducting the User Study. Green boxes indicate tasks to be addressed.](image_url)
1.3.1 MILESTONE 1A: Implement Experimental Platform

Here we present the main tasks for developing a web-based experimental platform that allows us to design and evaluate various versions of adaptive MMDs, with functionalities that include collecting a variety of users characteristics via on-line tests, displaying MMDs interactively, tracking how users are processing them, and providing adaptive support based on where the user is looking (see Figure 3A):

- **Acquire the MMD dataset**: A suitable set of MMDs are needed for testing our proposed adaptation mechanisms. Challenges include locating or producing a dataset that has identified the references, and also specifies the links between the text and the visualization for each MMD. Specifying the links entails generating an underlying representation of the data points contained in each of visualization, and specifying which subset of data points are associated to each reference. As a starting point, we have currently located a dataset of 40 MMDs from Kong et al. [2014] which has all of the references identified and the links fully specified. For further details on how we plan to work with this dataset see Section 3.1.

- **Reliably Automate Study Procedure**: The experimental platform should allow for experiments to be run in a streamlined and reliable way so that they can be easily administered even by non-specialized experimenters (e.g., non-computer science RA’s). Previous studies carried out in our lab have been cumbersome to administer due to complexity from needing to run several software applications sequentially (e.g., a different application for each user characteristic test), and due to the brittleness of our existing experimental platform (e.g., frequent crashes and inability to resuming where a crash left off). Identifying crashes in the software has been difficult because the components of the existing experimental platform were implemented using multiple programming languages, code wrappers, and development environments. We are currently developing a new and more robust experimental platform that has limited the number of programming languages to two (Python and JavaScript). The new platform will also streamline many of the user characteristic tests in order to reduce the number of separate software applications that have to be loaded by the experimenter during the study.

- **Set Up Connection with Eye Tracker**: The experimental platform will connect to an eye tracker in order to process a user’s gaze data in real-time. For instance, this will allow us to detect when the user is looking at a reference’s text in the MMD. Even though current software (e.g., Tobii Studio) processes gaze data in real-time, its primary function is for offline analysis of gaze data and the software does not provide us access to this gaze information in real-time. Therefore, we are currently working with an undergraduate RA to develop custom software to communicate directly with the eye tracker hardware via an API. Communicating with an eye tracker API is not necessarily a difficult task, however the challenge is to ensure that a real-time connection is maintained within the layers of software in our experimental platform so that adaptations will not be delayed.

- **Implement Intervention Mechanism**: This mechanism will manage the delivery of the adaptive interventions and requires the following functionality:
  - Capability to draw/animate a highlighting intervention on visualizations embedded in an MMD
- Capability to fade/remove the highlighting intervention off of the MMD
- Ability to configure interventions based on properties such as: type of intervention, location of intervention, intervention trigger time, etc.

The intervention mechanism will likely be implemented using the D3 JavaScript library since it provides suitable methods for animating and drawing over the interface.

**Milestone 1A Outcomes:** The outcome for this stage of work will be an experimental platform serving the following purposes:

- Allow for us to rapidly prototype various types of adaptive highlighting interventions
- Ability to run pilot studies for preliminary evaluation of intervention properties and study configurations
- Ability to conduct user studies with all of the functionalities mentioned above

### 1.3.2 MILESTONE 1B: Design and Conduct User Study

During this stage of work, we will address RQ1 and RQ2 by designing and conducting a user study (*Study 1*) using the experimental platform developed in the previous subsection. Figure 3B provides a high level overview of this research. We have identified several important tasks and challenges that will have to be addressed in order to complete the design of Study 1. These tasks are presented next, followed by a discussion of the Milestone outcomes:

- **Select Appropriate Set of User Characteristics:** Our plan is to choose user characteristics among those that have already been shown to impact visualization effectiveness (see Section 2.1), and possibly including others that are especially relevant to processing MMDs with visualizations (e.g. reading proficiency). However, the set of user characteristics that has already been shown to impact visualization processing is quite large (more than a dozen). In order to keep the length of the study manageable, the challenge will be in deciding which of these user characteristics to administer for Study 1. Refer to Section 4.1 for details on the user characteristics we are considering.

- **Determine Study Tasks:** The main criteria for guiding the selection of the study tasks will be to find tasks for which we can define clear and objective measures of performance (i.e., dependent variables in the study). For details on the different types of performance measures and tasks we are considering refer to Section 4.2.

- **Select Suitable Intervention Properties:** Some of the properties include, for instance, the type of highlighting provided (e.g., bolding, arrows), and if/for how long the delivery of the intervention should be delayed once a user has looked at a reference’s text (i.e., trigger time). Defining a reasonable starting set of highlighting types is challenging because there are many forms of highlighting to choose from, and they may all not be suitable. For instance, highlighting with *color change* could be misleading or confusing since color is often already used to encode information in the visualizations. Determining suitable trigger-time(s) is also challenging because we do not want to select values that are disruptive to the user (e.g., by providing the adaption
too soon or too late). For details on how we plan to produce a starting set of values for the intervention properties see Section 4.3.

- **Define Reference Complexity:** One factor that might influence the effectiveness of providing adaptive interventions is the complexity of the reference that the user is processing. For instance, complexity varies between a reference that identifies only one item in the visualization vs. a reference that compares several items in the visualization at once. Hence, we identified a set of 8 properties that could influence the complexity of a reference. We provide a formal way to define this complexity in a precise manner, so that reference complexity can be controlled for in our user study. For details on how we define reference complexity see Section 4.4.

- **Evaluate Other Possible Factors:** There might be other factors that need to be considered (e.g., length of the MMD, font size of the text, placement of the visualization) that we have not yet fully considered. Therefore, as we carry out the work in Milestone 1, we will plan to evaluate whether other possible study factors will merit inclusion in the design of the user study.

**Milestone 1B Outcomes:** The outcomes for this stage of work will be the answers to research questions RQ1 and RQ2. Namely, we will determine whether the evaluated highlighting interventions were effective at improving user experience, and if certain interventions are more effective depending on a user’s individual differences. The obtained outcomes will dictate how we proceed with our research into the next Milestone:

- **Outcome 0 – We were not able to detect any improvement in user experience for any adaptive interventions.**
  - This outcome could still be a contribution if we can ascertain reasons for why the adaptive interventions were not useful which could help define future steps for this research. First, we can check what the users had to say. We will ensure that post-study questionnaires are carefully constructed so that in the event that we reach this outcome, we will have data that may provide meaningful explanations from user responses. We may also consider conducting post-study interviews to the same effect. Second, we can analyze eye tracking data collected from the study to identify how users were processing the MMDs, and potentially why the interventions were not effective (e.g., because users were not looking at them). These findings could then be used to improve and fix our current interventions (e.g., by experimenting on how to make the interventions more visible). In addition, by analyzing users’ gaze patterns we could identify criteria for designing new types of interventions. For instance, if we found that users spent a significant portion of time processing the visualization’s axis-labels, new interventions could be devised with the aim of providing support targeted around this element of the visualization. Based on what we find from the post-study questionnaires/interviews and eye tracking analysis, we could then consider running a new study if there is sufficient ground for it. Alternatively, as stated above, the contribution would be compiling a set of guidelines for future work.
• **Outcome I** – We were able to detect an improvement in user experience, which would be a major contribution of this work; however we were not able to detect that individual differences played a role. Possible options to address this outcome include:
  o As described above, perform an analysis of the eye tracking data in order to see if we can detect differences in processing behaviors according to individual differences. Based on the results of this analysis, the contribution could be providing guidelines on how to improve our current interventions or for designing new interventions to support the differences in processing that we identified. First, since we already know that our interventions were useful, we could attempt to further improve them by identifying new properties of the interventions to experiment with. For instance, if we identified that two groups of users (e.g., users with high vs. low perceptual speed) had equal benefit from an intervention but processed it differently (e.g., low perceptual speed users spent more time looking at the non-relevant bars de-emphasized in the visualization), we could then consider experimenting with various intensities of the de-emphasized bars to see if there are optimal values tailored to each group of users. Second, we could identify guidelines for designing new types of interventions to support users according to their individual differences. To provide a concrete example, Toker et al. [2013] analyzed gaze data using a similar approach, and found that while working with bar graph and radar graph visualizations, users with low perceptual speed required significantly more time to process the visualization’s legend. This finding provided evidence toward where in the visualization new types of adaptive support could be targeted in order to support users based on their individual differences.
  o Consider evaluating additional user characteristics that would be derived post-hoc and see if they mediate the effectiveness of any of the adaptive interventions. For instance, modeling *learning curves* according to each user’s performance as described in Lallé et al. [2015].

• **Outcome II** – We detected that individual differences play a role in intervention effectiveness.
  o We will report our findings, and similar as above, we will analyze gaze data and questionnaire data to provide explanations of our results. We will then proceed to Milestone 2, discussed next.

---

**Figure 4.** High level overview of Milestone 2. Green boxes indicate tasks or challenges to be addressed.
1.3.3 MILESTONE 2: Predicting User Characteristics in Real-time

The purpose of Milestone 2 is to address RQ3. This work (see Figure 4) entails conducting machine learning experiments to evaluate the feasibility of predicting user characteristics from the eye tracking data collected in Milestone 1B. The motivation of this work is to offer a non-intrusive method to predict in real-time user characteristics found to be relevant in Milestone 1B, so that more personalized interventions can be provided to users as they process MMDs. Here we present an overview of the main tasks for Milestone 2, followed by a brief discussion of the Milestone’s outcomes:

- **Process User Characteristics Data:** Relevant user characteristics data collected in Study 1 will be processed in order to generate meaningful classification targets (i.e., the things we want to predict with our machine learning models). Most user characteristics are continuous variables, and since we don’t expect to have a dataset with thousands of users, predicting continuous variables with smaller datasets is difficult. Therefore, we will address this problem by transforming the user characteristics from continuous to discrete values. The details on how we will transform the user characteristics data remains an open question. For instance, past work has used the median split (e.g., high/low), however we may opt to produce more fine-grained targets (e.g., high/med/low). We plan on examining the distribution of values for each relevant user characteristic measured in Milestone 1B, to select suitable targets.

- **Process Eye Tracking Data:** Raw eye tracking data from Study 1 will be processed in order to generate suitable input features (i.e., the predictors) for machine learning. Eye tracking data will be processed with the Eye Tracking and Data Analysis Toolkit (EMDAT). EMDAT is open source software used to generate dozens of aggregate gaze features from raw eye gaze data. One challenge will be selecting which subset of gaze features to utilize for the machine learning experiments in order to avoid the curse of dimensionality (i.e., with finite training examples, prediction power decreases as the number of features/dimensions increases [Hughes 1968]). We provide further details on generating gaze features in Section 5.1.

- **Train and evaluate machine learning models:** Configuring machine learning models is challenging since there are many different algorithms available in standard ML suites (e.g., logistic regression, random forest, naïve Bayes) and there are many possible parameter settings for each of the different algorithms. Furthermore, there are more complex model configuration choices on how we select parameters for validating and segmenting the gaze features that will also have to be resolved. Hence, our task will consist of running machine learning experiments in order to identify which model configurations achieve the best accuracies for predicting user characteristics.

---

1. https://github.com/ATUAV/EMDAT
**Milestone 2 Outcome:** The outcome for this stage of work will be the answer to RQ3, namely, to what extent can relevant user characteristics be predicted in real-time from eye tracking data:

- **Outcome 0** – We were not able to predict relevant user characteristics from eye tracking:
  - This outcome could still be a contribution because it indicates that unlike previous work that has shown the feasibility of making predictions using similar techniques with visualizations only (see Related Work Section 2.3), this approach does not extend well to interactions containing visualizations and paragraphs of text.

- **Outcome I** – We were successfully able to predict relevant user characteristics from eye tracking:
  - This result provides important grounds for future work, namely, the means for a user-adaptive MMD system to predict individual differences in real-time in order to guide the selection of the interventions accordingly.

**1.4 Contributions**

This work is expected to help advance the research in information visualization, multi-modal interactions, user modeling, user-adaptive interactions, and eye tracking in various ways. The research contributions of this work include:

- It is the first work to explore providing adaptive highlighting interventions to Multi-Modal Documents that include text and visualizations, delivered based on where the user looking.
- Quantitative results on the effectiveness of various adaptive highlighting techniques, or explanations for lack of effects that can inform future work.
- Quantitative results on the influence that various user differences have on the effectiveness of adaptive highlighting techniques, or explanations for lack of effect that can inform future work.
- A set of definitions that allows for a precise definition of MMD reference complexity
- New understanding on the feasibility of using eye gaze data and machine learning techniques to predict user characteristics in real-time as users interact with MMDs.

In addition, this work aims to make several practical contributions:

- An MMD dataset suitable for running future user studies. The dataset also contains coded metadata for the reference properties (e.g., reference complexity, links to referenced data-points in the visualization, visualization type, etc.). This dataset will be made available to the visualization community.
- A new dataset containing eye gaze data collected from users interacting with MMDs, which can be made available to the visualization and machine learning community for carrying out future analyses.
- Software of our experimental framework that can be made available on GitHub for other researchers. The software could be further extended to suit the needs of other studies wishing to use real-time eye tracking and adaptive interventions in similar or other domains (e.g., different types of MMDs, intelligent tutoring systems,).
The rest of this proposal is structured as follows. First we review related work in Section 2. Next, Section 3 presents work on selecting a MMD dataset (Milestone 1A). After that, Section 4 presents work and challenges for designing and conducting a user study. Then, Section 5 presents the proposed details on processing eye tracking data to predicting user characteristics for Milestone 2. Finally, Appendix A provides a research timeline as well as details on previous and proposed research papers.
2. Related Work

The related work we cover in this section focuses primarily on the field of information visualization research because i) the type of Multi-Modal Documents (MMDs) our work is concerned with all contain visualizations, and can thus be seen as a logical extension of visualization research; and ii) it will be the visualizations themselves that will be the primary target for providing adaptive highlighting interventions. First, we cover work on how various individual differences have been shown to impact visualization effectiveness. Next, we examine several user-adaptive visualization systems that have been previously designed and evaluated. Finally, we cover work on using machine learning to predict individual differences from eye tracking data.

2.1 Individual Differences in Information Visualization

Despite the fact that most visualizations are still designed and deployed following a one-size-fits all approach, a significant body of work has established that the effectiveness of visualizations can depend on individual differences of the user. Such differences (also referred to as user characteristics) can encompass a wide range of factors including: cognitive abilities, personality traits, affective states, skill or expertise, and preferences. Showing that visualization effectiveness can depend on individual user differences is important because it motivates the need for user-adaptive visualizations, which are visualizations that can customize the interaction to support users according to their individual needs. In other words, by uncovering which user differences matter, visualization systems can then be designed to adapt to these identified factors. Table 1 provides a list of individual differences consisting of cognitive abilities and personality traits that have previously been shown to impact InfoVis effectiveness. Cognitive abilities are “aspects of mental functioning, such as memorizing and remembering; inhibiting and focusing attention; speed of information processing; and spatial and causal reasoning” [Robinson 2012]. Personality traits “describe a person’s dispositional and distinctive pattern of thoughts, feelings, and behavior across various situations” [Heinström 2012]. Table 1 is useful because it can help guide the selection of which user characteristics we may want to consider in our research. Details surrounding the findings for the user characteristics presented in Table 1 will be the focus of rest of this subsection.

<table>
<thead>
<tr>
<th>User Characteristic</th>
<th>Definition</th>
<th>Test Name</th>
<th>Relevant research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptual Speed</td>
<td>A measure of speed when performing simple perceptual tasks (e.g., finding figures or making comparison between letters, numbers, objects, pictures, or patterns) [Salthouse 2000].</td>
<td>Identical Figures Test (P-3) [Ekstrom 1976].</td>
<td>[Allen 2000], [Velez 2005], [Conati &amp; Mclaren 2008], [Toker et al. 2012], [Careini et al. 2014], [Conati et al. 2014]</td>
</tr>
</tbody>
</table>
### Visual Working Memory

Part of the working memory responsible for temporary storage and manipulation of visual and spatial information [Logie 1995].

- Fukuda & Vogel colored squares test [Fukuda & Vogel 2009].
- [Toker et al. 2012], [Careini et al. 2014], [Conati et al. 2014]

### Locus of Control

Measure of the degree to which individuals perceive outcomes as either a result from their own behavior, or from forces that are external to themselves.

- 29-item locus of control scale [Rotter 1966].

### Extraversion

Extraversion defines the degree to which a person is open-minded, action-oriented and seeks the society of others.

- Mini-IPIP scales of Big Five factors of personality [Donnellan et al. 2006].
- [Green & Fisher 2010]

### Neuroticism

Neuroticism is distinguished by negativity and a propensity to be moody.

- Mini-IPIP scales of Big Five factors of personality [Donnellan et al. 2006].
- [Green & Fisher 2010]

### Spatial Orientation

The ability to identify the spatial rotation of 3-dimensional objects.

- Cube Comparison Test (S-2) [Ekstrom 1976].
- [Velez et al. 2005]

### Spatial Visualization

The ability to manipulate or transform the image of spatial patterns into other arrangements.

- Paper Folding Test (VZ-2) [Ekstrom 1976].
- [Velez et al. 2005], [Conati & McLaren 2008]

### Disembedding

The ability to hold a given visual percept or configuration in mind so as to disembed it from other well defined perceptual material.

- Hidden Patterns Test (CF-2) [Ekstrom 1976].
- [Velez et al. 2005], [Conati & McLaren 2008]*

### Visual Memory

The ability to remember the configuration, location, and orientation of figural material.

- Shape Memory Test (MV-1) [Ekstrom 1976].
- [Velez et al. 2005], [Conati & McLaren 2008]

### Need for Cognition

Measures the extent to which individuals are inclined towards effortful cognitive activities.

- 18-Item Need for Cognition Scale [Cacioppo et al. 1984]
- [Conati & McLaren 2008]

### Learning Style

Preferences for the manner in which information is received and learned.

- [Felder & Silverman 1988]
- [Conati & McLaren 2008]

### Associative Memory

The ability to recall one part of a previously learned but otherwise unrelated pair of items when the other part of the pair is presented. [Carroll 1974]

- Associative Memory Test (MA-1) [Ekstrom 1976].
- [Chen 2000]

| **Table 1.** List of cognitive abilities and personality traits that have been previously investigated in relation to visualization effectiveness. | * indicates that no significant results were identified for that given user characteristic. |

In Allen [2000], the author examined if perceptual speed plays a role in the effectiveness of 4 different configurations of a visualization used to explore bibliographic entries and keywords. The visualization consisted of: i) either a single window or multiple windows used to visualize keywords, descriptions, and metadata of bibliographic entries; and ii) either the presence or not of a world-map used to visualize the bibliographic keywords. Users were asked to select articles for a hypothetical research paper, and the
selections were scored based on the relevance of the chosen articles. The authors found that users with low perceptual speed scored better with the world map or single-window display, whereas users with high perceptual speed scored better when no world-map was displayed or when working with a multi-window visualization.

In Velez et al. [2005], the authors explored the link between three spatial abilities (spatial orientation, spatial visualization, and disembedding) and two cognitive factors (visual memory and perceptual speed). They conducted a study measuring accuracy and response time for tasks requiring user to identify a 3D object according to its orthogonal projections. They found that all of the spatial abilities positively correlated with task accuracy, and that both cognitive abilities negatively correlated with response time.

Conati & Maclaren [2008] conducted a user study to investigate how several individual differences relate to the compared effectiveness of two different visualizations. The individual differences examined included: perceptual speed, visual memory, spatial visualization, disembedding, need for cognition, and learning style. The visualizations were used to compare how various indicators (e.g., economic, environmental, social health) were subject to change in a given geographic region based on various scenarios. In the study, users interacted with either a radar chart visualization to compare changes in a given scenario based on 9 high-level indicators; or a Multiscale Dimension Visualizer (MDV) [Williams & Munzner 2004] which facilitated comparisons across 294 low-level indicators. Results showed that users with low perceptual speed were significantly more accurate with the MDV, whereas users with high perceptual speed were more accurate with the radar chart. The study also administered several visualization tasks (e.g., filter, sort, characterize distribution, etc.), and varied the quantity of scenarios to compare (1 scenario vs 2 scenarios). Further results indicated that the optimal visualization for certain task and scenario configurations were also dependent on a user’s visual memory, perceptual speed, spatial visualization, need for cognition, or learning style (see paper for further details). No results were found for the individual difference: disembedding.

Toker et al. [2012] ran a user study to investigate if perceptual speed, verbal working memory, and visual working memory influenced the compared effectiveness of bar graphs and radar graphs. They found that users were always faster with bar graphs regardless of individual differences. However, users with low perceptual speed performed significantly worse with radar graphs. In addition, results indicated that users with high visual working memory had a stronger preference for radar graphs, and users with low verbal working memory found bar graphs easier to use.

Carenini et al. [2014], ran a study evaluating simple and complex tasks using grouped bar charts. Their findings indicated that users with low perceptual speed, users with low verbal working memory, and users with low visual working memory, all performed significantly worse when completing more difficult tasks, indicating that harder visualization tasks leads to greater strain on each of these three cognitive abilities.

In Conati et al. [2014], the authors reported on a study involving an interactive stacked bar chart visualization used for decision making (ValueChart [Carenini & Lloyd 2004]), and found that perceptual speed, verbal working memory, visual working memory, all played a role in how effective the
visualization was depending on the type of task the user was doing. Furthermore, they also found that users with low visual working memory were significantly faster with a horizontal layout of the visualization, whereas users with high visual working memory were significantly more accurate with a vertical layout of the visualization. These findings are important because previous work linking visual working memory to performance with visualization tasks found that users with low visual working memory were at a disadvantage.

Green & Fisher [2010] examined the effect of three personality traits (locus of control, extraversion, and neuroticism) on tasks carried out with two different visualization interfaces used for displaying hierarchical genomic information (Gvis and MapViewer). In terms of task completion time, users with internal locus of control were faster with GVis compared to users with external locus of control. In addition, results showed that users with higher levels of neuroticism or extraversion were significantly faster with both visualizations.

Ziemkiewicz et al. [2011] examined various visualizations of phylogenetic tree data by using 4 different layout styles: Basic Tree; Bordered Tree; Indented Boxes; and Nested Boxes. They found that users with internal locus of control were much slower at correctly solving tasks with the Nested Box view compared to the basic Tree View, whereas users with external locus of control were just as proficient with both views.

Chen [2000] found that users with higher associative memory achieved higher performance scores when using a spatial interface for visualizing the semantic structure of research articles.

Ottley et al. [2015] conducted a user study comparing search tasks using two popular hierarchical visualizations: Indented Tree and a Dendrogram. Their results indicated a significant interaction effect between visualization type and locus of control on correct response time. In particular, users with external locus of control performed better with the Indented Tree visualization compared to users with internal locus, whereas users with internal locus performed better with the Dendrogram compared to users with external locus of control.
2.2 User-Adaptive Visualization Systems

Here we provide a few examples of existing user-adaptive visualization systems have shown that the effectiveness of the visualization system can be improved when providing useful adaptation tailored to the needs of the individual differences which provide motivation toward designing user-adaptive MMDs for our research.

Gotz & Wen [2009] evaluated an information visualization system using a technique they developed called: Behavior Driven Visualization Recommendation (BDVR). The approach entails identifying various behavioral patterns during interaction with the system, and then providing a suitable recommendation to the user. The visualization system is intended to facilitate the discovery of insights from large datasets by allowing the data to be visualized using a large quantity of different types of information visualizations (e.g., bar charts, line charts, radial graphs, maps, networks, trees, etc.), see Figure 5. A user begins by selecting a visualization of their choice to carry out search and comparison tasks with a large dataset, and then based on the various sequences of actions a user makes while working with that visualization (i.e., a pattern), the system will respond by recommending a more suitable visualization based on the detected pattern. Examples of patterns include: a Swap pattern (users repeatedly rearrange the order of the data); and a Scan pattern (users perform multiple inspections over a series of visual objects). Based on a user study comparing the visualization system with BDVR enabled versus not, Gotz & Wen [2009] reported that the system was able to detect patterns in 41 out of the 60 trials with BVDR enabled, and in 36 of these 41 cases users accepted the suggested recommendation for an alternative visualization. Most importantly though, they reported a statistically significant reduction in task completion time of 21%, and a reduction in error rate of 81% when BVDR was enabled. These results evidence that providing real-time visualization recommendations based on individual user behavior is worthwhile.
Grawemeyer [2006] created and evaluated an adaptive system that would recommend a more suitable visualization or limit the choice of which visualizations could be selected when asking users to select an appropriate visual representation for a series of database queries (see Figure 6). Similar to Gotz & Wen [2009], the user model consisted of tracking users’ behavior during interaction with the system. In particular, the system would track the user’s evolving knowledge of external representations, and recommendations for visualizations suitable to the user’s current knowledge were provided at the outset of each new task. Results from the user study conducted to evaluate the system showed that the adaptive version of the system led to improved user performance compared to the non-adaptive version.

The two user-adaptive visualization systems we just presented both provided adaptive support by recommended more suitable visualizations. In terms of which individual differences they tracked in the user model, they adapted based on either user goals or user knowledge. In contrast with our research, we plan on providing adaptations within the current visualization in the MMD (as opposed recommending a different visualization). Furthermore, we are planning on adapting according to a user’s reading behavior (as opposed to their goals or knowledge).

### 2.3 Predicting User Characteristics from Eye Tracking Data

The purpose of this section is to provide examples of work establishing the feasibility of predicting various user characteristics using eye tracking. Establishing this feasibility is important because it is the main objective of Milestone 2 (see Section 1.3.3). First, we present work on predicting user characteristics from eye tracking data generated from information visualizations tasks. Second, we present work doing the same in other non-visualization domains.

In the domain on InfoVis, Toker et al. [2014] showed that eye tracking could be used to predict a user’s skill acquisition state during bar graph visualization tasks. Both Steichen et al. [2014] and Gingerich & Conati [2015] showed that perceptual speed, verbal working memory, visual working memory could be predicted using eye tracking data from tasks with radar charts and bar charts. Lallé et al. [2015] showed that eye tracking could be used to predict a user’s learning curve as users completed a series of tasks.
with an interactive bar chart. Lallé et al. [2016] also showed that eye tracking could be used to predict instances of user confusion while interacting with an interactive stacked bar chart visualization.

Work outside of InfoVis has also been carried out to predict user characteristics from eye tracking data. For instance, Bixler et al. [2015] predicted mind wandering during reading tasks. Martinez-Gomez & Aizawa [2014] predicted language skill and understanding level during reading tasks. Kardan & Conati [2013] predicted users either having high learning gain versus low learning gain while using an interactive simulation for learning about constraint satisfaction problems in Lastly, Copeland et al. [2015] predicted reading comprehension during reading tasks.
3. MILESTONE 1A: Implement Experimental Platform

The primary objective of this Milestone is to build a web-based experimental platform in order to carry out the design and evaluation of adaptive MMDs. Functionalities will include: collecting user characteristics data from on-line tests, displaying MMDs and tracking how users process them, as well as providing adaptive interventions based on where the user is looking. We present the tasks for building the experimental platform in Figure 7. The first three tasks are related to creating the software for the experimental platform (i.e., automating the study procedure, setting up the eye tracker connection, and implementing the intervention mechanism.) Details regarding the current description and challenges of these three software tasks are presented in the Introduction (Section 1.3.1). Here, we will address the task of acquiring a Multi-Modal Document (MMD) dataset by presenting one that we have selected for our research. We will also discuss additional work needed to make the dataset suitable to our purposes as part of this task.

![Figure 7. Tasks for implementing the experimental platform.](image)

3.1 Acquire MMD Dataset

A suitable dataset of MMDs will be needed for users to interact with while carrying out study tasks. The challenges with acquiring a dataset include locating the set of relevant documents themselves, then identifying the references in the MMD, namely locating reference sentences in the MMD text and the corresponding data points within an in the accompanying visualization targeted by each reference. Addressing these challenges, we came across work from Kong et al. that have already generated a suitable starting dataset (Kong dataset from now on) of 40 MMDs with 182 references [Kong et al. 2014]. The Kong dataset was derived from real-world sources including Pew Research (a social science research group: www.pewresearch.org), The Guardian (a daily newspaper: www.theguardian.com), and The Economist (a weekly news magazine: www.economist.com). Each MMD in the Kong dataset consists of one paragraph of text and one bar graph visualization. The references in each MMD were identified by Kong et al., along with coding indicating which data points in each visualization correspond to each...
Reference text. The authors also created a basic interface used to explore the references in the MMDs. Reference text is selected in the interface via mouse input by the user. Upon selection, red highlighting is applied over both the selected text and the corresponding elements in the visualization (see Figure 8).

One possible limitation of the Kong dataset is that the majority of its MMDs include only bar-chart based visualizations. However, we argue that focusing on this type of visualizations is adequate for our purposes. Reducing the scope of the targeted visualizations is necessary to keep the complexity of the research manageable for a PhD thesis. The research outcomes however, would still be significant, because given the familiarity of bar-chart based visualizations, they are widely used in MMDs from sources that target a general public (e.g., the Economist).

Figure 8. Example MMD in the Kong dataset, displayed using their interface. Here, one reference has been selected in the text, and the nine corresponding bars in the visualization have been highlighted accordingly.

Despite the obvious value of the Kong dataset for our research, there are some issues with the format of the MMDs that will need to be addressed as part of this research. For instance, each MMD in the Kong dataset consists of “snippets” of larger documents, where each snippet includes exactly 1 paragraph of text and 1 accompanying visualization (see Figure 9 for an example). The reason for this format is in the original goal of the research behind the Kong dataset [Kong et al. 2014], namely proposing a novel approach to automate the extraction of references in MMDs by utilizing crowdsourcing and clustering techniques. Breaking longer MMDs into single-paragraph-single-viz snippets was necessary to facilitate coding via crowdsourcing [Kong et al. 2014].

2 The interface, along with the full Kong dataset can be accessed at: www.cs.ubc.ca/~lalles/TextGraphVis/Site.html
One issue with the single-paragraph-single-viz format is that it generates **fragmentation**, occurring when a sentence in a Kong MMD is difficult to comprehend because it refers to information from paragraphs in the original documents that were not included in the present Kong MMD. Figure 10 provides an instance of fragmentation (underlined in red). We plan to address this issue by either removing the fragmented text when possible, or by extending the fragmented MMD to include the relevant paragraphs. The latter will entail either combing existing snippets when the necessary paragraphs are in the Kong dataset, or retrieving the missing text from the original documents, to which we have access. Notice that the option of extending fragmented MMDs is viable because we are not constrained to maintaining the single-text-single-viz format for our purposes.

As a matter of fact, even when MMDs in the Kong dataset are not fragmented, we may want to combine some of those that come from the same original document in order to have more realistic MMDs for our proposed studies. We found that 75% of the original 40 Kong MMDs are combinable, yielding 10 multi-paragraph-multi-viz MMDs and 10 single-paragraph-single-viz. Whether or not to administer MMDs that contain more than one paragraph of text or more than one visualization is an open question that we will investigate as part of the design on our studies. Using longer MMDs has higher ecological validity since most real-world MMDs contain multiple paragraphs of text and possibly multiple visualizations. In contrast, using only the basic uncombined Kong MMDs may be more suitable for our preliminary work since it reduces the complexity of the dataset and study design. Finally, additional work will be required to ensure the Kong dataset is balanced for our user studies. For example, if the type of bar graph becomes a controlled factor in the study (see Section 4.4), then balancing the Kong dataset will likely require producing additional MMDs to ensure that there is an equal representation of each type of graph.
It matters because people who feel they belong where they live are also more likely to be happier. Some 82% of those aged 50 and over who agreed that they belonged to their neighbourhood were satisfied with their lives. For those who did not feel they belonged, the rate was 53%.

Figure 10. Example of fragmentation in a Kong MMD, since it is unclear what is meant by ‘It matters’ (underlined in red).
4. **MILESTONE 1B: Design and Conduct User Study**

In this chapter we will present our plans for designing and conducting a user study in order to address research questions RQ1 and RQ2.

**RQ1:** To what extent does providing adaptive highlighting interventions, triggered when the user looks at a reference’s text in an MMD, improve the effectiveness of the MMDs?

**RQ2:** To what extent does the effectiveness of the highlighting interventions in RQ1 depend on the individual differences of the user?

First, we will present several tasks that have to be completed as part of the design of the user study (green boxes in Figure 11). Next, we will discuss how we plan to analyze data generated by the study, and finally we will present a discussion on the possible study outcomes and contributions.

![Figure 11. Overview of the user study indicating tasks that will have to be addressed.]

### 4.1 Select Appropriate Set of User Characteristics

We are currently considering the following characteristics:

- The cognitive abilities Perceptual Speed [Ekstrom 1976], Visual Working Memory [Fukuda & Vogel 2009], and Verbal Working memory [Turner 1989] because, as we illustrated in the related work section, it has been repeatedly shown that the effectiveness of visualizations can depend on these abilities. These visualizations included single bar charts, grouped bar charts, stacked bar charts, radar charts, tree maps, and word maps. Our aim is to include the abovementioned cognitive abilities to see if these previous findings will extend to our research, since our MMDs contain similar visualizations (i.e., single/grouped/stacked bar charts).
- Personality traits in the Five Factor model [Donnellan et al. 2006], and Locus of Control [Rotter 1966], since as discussed in the related work section, they too have been shown to play a role in the
effectiveness of visualizing hierarchical data (e.g., trees, tree maps, nested views, and dendrograms). However, these personality traits have not yet been linked to bar graph visualizations. Including these personality traits in our study could still be worth considering since MMDs require users to process multiple information sources beyond just a visualization (i.e., visualization, text, and references).

- Reading skill (e.g., reading comprehension, language comprehension), since processing references in MMDs includes reading one or many paragraphs of text.
- Visualization Literacy with bar charts [Boy et al. 2014], since processing references in MMDs also includes processing bar chart visualizations. Since this measure is relatively new, there is little related work assessing it in other studies. However, it is the most promising measure we are currently considering for our study to assess a user’s overall skill in processing bar charts.

Ultimately, the main concern with selecting the set of user characteristics is the tradeoff between ensuring a comprehensive set, while not taking up excessive study time to administer the tests. One solution to accommodate a larger set of user characteristics include running a longer study session (e.g., 3 hours) with a break in the middle. Another solution would be to automate the tests in such a way that users could complete them remotely before the study, as it has been done for a different project in our research group.

4.2 Select Suitable Intervention Properties

In this section, we present 3 properties which describe the adaptive interventions on MMD visualizations that we plan on evaluating in our user study. The challenge of this work rests in determining values for each property so that the adaptive interventions will serve to improve user experience with MMDs as opposed to being disruptive. For each property, we will provide our rationale on identifying which values we are considering using, along with any open questions. We plan on leveraging our experimental platform to rapidly prototype various intervention configurations (i.e., different combinations of properties and values), and then evaluating them via pilot tests. Our aim is to identify a minimal and optimal starting set of values for the intervention properties in order to prevent the complexity of the study from growing too large. It should be noted that here we discuss only interventions on the visualization components of MMDs. Investigating highlighting techniques in the MMD beyond the visualization will likely be left for future work. Options would include also highlighting the reference’s text, or visually linking (e.g., with a line) the reference’s text with the items highlighted in the visualization (as described for instance in [Steinberger et al. 2011]).

4.2.1 Type of Intervention Highlighting

Here we present the types of highlighting we are considering for our interventions. As we discussed in related work, Carenini et al. [2014], evaluated four different highlighting interventions (Figure 12) designed to guide users’ attention to relevant parts of a bar graph visualization to help them answer specific questions on the visualized dataset. All four interventions were shown to improve user experience (i.e., time on task and subjective ratings) compared to providing none. Specifically, *De-Emphasis* was found to be the most effective, and *Bolding* and *Connected Arrows* were tied for second best. These findings are relevant to our work because the bars highlighted in the visualization were the
ones referred to by the visualization task, which is exactly what we want to have for guiding MMD processing. The main difference with our work is that in Carennini et al. [2014], highlights were used to help users answer specific questions, a scenario suitable for the purposes of that study but not very realistic. We want to ascertain if/how their results generalize to our settings, hence we plan on selecting highlighting types from the 4 presented in Carennini et al. [2014]. How many we will select remains an open question. For instance, in order to minimize study complexity we could select only two highlighting types that performed at best in Carennini et al. [2014] (e.g., De-emphasis & Connected Arrows). Alternatively, we could opt to test all 4 highlighting types since we cannot be certain to what extent the results from Carennini et al. [2014] will generalize to MMDs.

![Types of highlighting interventions evaluated in Carennini et al. [2014].](image)

4.2.2 Timing of Intervention Triggering

Here, we present options for when to provide an intervention to the user. Determining when to trigger an adaptive intervention is a noteworthy challenge because we want to offer users support with the task at hand, while minimizing the cost of interruption. In Carennini et al. [2014], they evaluated highlighting interventions under two delivery conditions as a first step toward addressing this challenge. The interventions were provided either: a) at the outset of each task (static condition); or b) triggered after some time had elapsed in order to simulate more realistic conditions where interventions are provided while the user is working on the task (dynamic condition, which can be potentially disruptive). Results showed that in the dynamic condition, the effectiveness of some interventions decreased, but they were all still better than providing no interventions at all. In our work with MMDs, we will be able to initiate adaptations in a more meaningful manner because adaptations will be provided when users look at reference’s text in the MMD. An open question however, is precisely when to trigger the adaptation once the user has looked at the reference’s text. For instance, we could experiment using three different trigger times according to how much of the reference’s text a user has read, e.g., 1) when the user looks at the beginning of the reference’s text; 2) when the user has read half the reference’s text; or 3) when the user has completed reading all of the reference’s text.
4.2.3 Timing of Intervention Removal

Here we present options on if/when interventions should be removed from the screen after they have been triggered. Since most MMDs contain more than one reference, we are faced with the challenge of handling situations where a user looks at multiple references as they process an MMD. Here, we present four options that we are considering to address this problem:

i. Allow for only one intervention to be displayed at a time:
   - Each intervention would be removed after a fixed amount of time has elapsed. New interventions would be triggered only if there are none on the screen, or
   - Each intervention would be replaced when a new one is triggered. However, this could become disruptive if several references appear close together in the text.

ii. Allow for multiple interventions displayed at a time:
   - Each intervention would be removed after a fixed amount of time has elapsed, or
   - Interventions would not be removed.

In the event that we allow for multiple interventions at once, we may have to include an additional intervention property to help distinguish interventions from one another (e.g., by using color or line style).

4.3 Determine Study Tasks & Dependent Measures

One key factor guiding the selection the tasks that users will perform in the study is to have tasks for which we can define clear and objective measures of performance to be used as dependent variables in the study. The candidate task types that we are considering include:

- **Comprehension-based** – Users will be given the general task of reading and understanding a set of MMDs, and after some allotted time they will be asked a set of questions designed to test their comprehension of relevant concepts discussed in the documents. This type of task is the most ecologically valid in reflecting how people process MMDs when they read them in everyday settings (e.g., reading an article of interest in The Economist).
- **Response-based** – Users will be given specific questions requiring them to search the MMD to find the correct response (e.g., retrieve the value for a specific variable, compare two given variables).
- **Verification based** – Users will be asked to specify whether a given MMD is correct. The idea is that some of the references could be altered so that the text incorrectly references the visualization in some way. It would then be the task of the user to examine the MMD and determine if the document’s references are correct or not.

Regardless of the task, in addition to objective measures of task performance we will also consider subjective measures of user satisfaction, e.g. perceived usability. We will collect these measures via post-questionnaires, which will be designed to provide us with detailed information on the reasons for the specific ratings generated. This information will be especially useful in case of negative study results (e.g., no effects of interventions in RQ1), in order to understand the reasons for the outcome and still be able to advance knowledge in user-adaptive visualizations via our lessons learned.
4.4 Define Reference Complexity

There are several factors that can influence the difficulty or complexity that a user may encounter when processing a reference in an MMD such as: the specificity of the reference, the visual task expressed in the reference, or the size of the dataset visualized. For example in Figure 13, the reference on the left has lower complexity since the text refers to only one bar in the visualization, the bar itself is easy to locate since the values can be matched (93%), and the whole visualization is relatively small with only 4 bars. The reference on the right in Figure 13 is considerably more complex since it refers to many bars in the visualization, the bars are harder to find since the names of the values are not provided in the reference’s text, and the visualization itself has considerably more bars which can be of two different colors. Reference complexity could have an impact on the relative effectiveness of the adaptive interventions we plan on evaluating. There is already an indication from Carenini et al. [2014] that the effectiveness of some highlighting interventions depended on the complexity of the visualization tasks being performed. Thus, it is important that we can control for reference complexity in our study.

![Less complex reference: “93% of Asian Americans describe members of their country of origin group as very hardworking”]{fig:less_complex}

![More complex reference: “Half or more in 13 of the 21 nations surveyed believe that most people can succeed if they are willing to work hard.”]{fig:more_complex}

Figure 13. Two references from the Kong dataset. Highlighting indicates which bars are being referenced respectively.

In order to describe the various properties of the MMDs that could impact complexity between references with consistency and precision, we will need a formal and quantitative approach, which does not yet currently exist. In this section, we present our work on identifying and defining a set of properties of the MMDs intended to capture the various dimensions of reference complexity. We will use this collective set of properties as a precise measure of reference complexity in our user study since it could have an impact on the relative effectiveness of the adaptive interventions we plan on evaluating. For instance, there is already an indication from Carenini et al. [2014] that some highlighting
interventions become significantly less effective when they are provided to users during visualization tasks with higher complexity.

We have already mentioned in the introduction that a reference in In a Multi-Modal Document is a set of words (typically a sentence) that specifies a visual task on an accompanying visualization embedded in the MMD. Here we provide a more formal definition of an MMD reference based on InfoVis constructs defined in Munzner [2014]:

“A mark is a basic graphical element in an image” Munzner [2014], and in the case of bar charts, each bar in the visualization is a mark.

A target is “some aspect of the data that is of interest to the user” Munzner [2014]. In the case of MMD references, the marks in the visualization specified by the visual task in the reference are the targets.

Based on these two definitions, a reference is a pair $<w, t>$ such that:

- $w$ is a set of words in the MMD text,
- $n$ are all the marks in a visualization $v$ in the MMD, and
- $t$ are the targets of $w$, where $t$ is a subset of $n$

According to the above definition of a reference, we identified three suitable categories$^3$ that we will consider to define properties of the MMD that may capture reference complexity: a) properties of the visualization; b) properties of the words in the reference’s text; and c) properties of the links between the reference and visualization (i.e., which targets in the visualization and how they are elicited in the text), expressed as visual tasks. Table 1 presents the initial set of properties that we identified as capturing various elements of reference complexity, grouped according to these three categories. These properties were identified by focusing only on references in the Kong dataset (see Section 3.1 for a description of this dataset). Practically speaking, since we will primarily use MMDs from the Kong dataset for our study, our aim is to produce a comprehensive enough set of properties to adequately describe the references in the Kong dataset. In the rest of this section, we provide definitions for each of the properties we identified, along with examples and statistics from the Kong dataset when possible.

---

$^3$ We are also considering a fourth category pertaining to the body of text in the MMD that the reference is situated in, e.g., number of paragraphs in the entire MMD, number of words in the entire MMD. Including this category in our definition of reference complexity is currently an open question.
### 4.4.1 Properties of the Visualization

We identified four properties of the visualization that can influence the complexity of a reference. One is visualization type. In the Kong dataset, all of the visualizations are bar charts. Within this dataset, we identified 6 different visualization types based on how the bar charts were configured, shown in Figure 14. We also provide statistics on quantity of each of these visualization types in the Kong dataset in Table 3, most of the visualizations in the Kong dataset are single or grouped bar charts.
The other two properties of the visualization we identified capture the size and complexity of the data to be represented. We leverage definitions from *Visualization Analysis & Design* [Munzner 2014] in order to describe these properties.

### 4.4.1.1 Levels of Primary and Secondary Key

Munzner [2104] employs a *data table* to describe the data shown in a visualization. A data table consists of rows and columns such that each row of data is represents an *item* in the dataset, each column represents an *attribute* defined for the items in the dataset, and each cell provides a *value* for a specific item/attribute pair.

For 1-dimensional data (which can be visualized for instance using single bar charts), we require a data table with two attributes to represent the data:

- **Primary Key attribute**: Categorical attribute that captures the distinct items in the dataset. When visualized with a bar chart, this attribute is used to separate the marks along the main axis of the visualization.
- **Value attribute**: Quantitative attribute that expresses the values of the items in the dataset. When visualized with a bar chart, the values are expressed with aligned vertical or horizontal spatial position.

The levels of the primary key equals the number of rows in the data table. An example of how this applies to a single bar chart in the Kong dataset is provided in Figure 15.

<table>
<thead>
<tr>
<th>Primary Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egypt</td>
<td>1,700,000</td>
</tr>
<tr>
<td>Iraq</td>
<td>15,804,557</td>
</tr>
<tr>
<td>Turkey</td>
<td>22,094,911</td>
</tr>
<tr>
<td>Regional</td>
<td>28,029,721</td>
</tr>
<tr>
<td>Lebanon</td>
<td>37,812,334</td>
</tr>
<tr>
<td>Jordan</td>
<td>40,221,893</td>
</tr>
</tbody>
</table>

**Figure 15.** Data table consisting of 2 columns, visualized using a single bar chart. There are 6 levels in the primary key.
For 2-dimensional data (which can be visualized for instance using grouped/stacked bar charts), we require a data table with a third attribute:

- **Secondary Key attribute**: Categorical attribute that captures a second dimension in the dataset. Values for distinct items in the dataset are provided by a <primary key, secondary key> pair. When visualized using a grouped/stacked bar chart, the levels of the secondary key can be encoded for instance using color.

With a 2-dimensional data table, the total number of values is equal to the levels of the primary key multiplied by the levels of the secondary key. Figure 16 provides an example, visualized using a grouped bar chart.

![Data Table](image)

**Data Table**

<table>
<thead>
<tr>
<th>Primary Key</th>
<th>Secondary Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol Use</td>
<td>Used Adderall</td>
<td>95.4</td>
</tr>
<tr>
<td>Alcohol Use</td>
<td>Did Not Use Adderall</td>
<td>63.0</td>
</tr>
<tr>
<td>Binge Alcohol Use</td>
<td>Used Adderall</td>
<td>89.5</td>
</tr>
<tr>
<td>Binge Alcohol Use</td>
<td>Did Not Use Adderall</td>
<td>41.4</td>
</tr>
<tr>
<td>Heavy Alcohol Use</td>
<td>Used Adderall</td>
<td>55.2</td>
</tr>
<tr>
<td>Heavy Alcohol Use</td>
<td>Did Not Use Adderall</td>
<td>15.6</td>
</tr>
</tbody>
</table>

![Visualization](image)

**Visualization**

Figure 16. Data table consisting of 3 columns. There are 3 levels for the primary key, 2 levels for the secondary key, yielding a total of 6 values. The data is visualized using a grouped bar chart.

Of the 40 visualizations in the Kong dataset, there are a total of 13 one-dimensional datasets (i.e., single bar charts), and a total of 27 two-dimensional datasets (grouped/stacked bar charts). We present statistics on the levels of the primary and secondary keys in Table 4.

<table>
<thead>
<tr>
<th>Kong Visualizations</th>
<th>Total</th>
<th># of Levels</th>
<th>Avg</th>
<th>Min</th>
<th>Max</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Dimensional datasets</td>
<td>13</td>
<td>Primary Key</td>
<td>7.4</td>
<td>3</td>
<td>13</td>
<td>4.2</td>
</tr>
<tr>
<td>2-Dimensional datasets</td>
<td>27</td>
<td>Primary Key</td>
<td>11.1</td>
<td>2</td>
<td>27</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary Key</td>
<td>2.5</td>
<td>2</td>
<td>4</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Table 4. Statistics on the levels of the primary and secondary keys for the 40 visualizations in the Kong dataset.
4.4.2 Properties of the Text

We identified three properties of the reference’s text that aim to capture varying difficulty in processing the textual elements of the reference.

4.4.2.1 Number of Words

A total of all the words contained in the reference’s text.

4.4.2.2 Number of Sentences

The total number of sentences that the reference’s text spans (typically 1, sometimes 2).

4.4.2.3 Readability Score

Difficulty of the reference’s text based on traditional quantitative reading based measures. For example, the Flesch–Kincaid readability test [Kinkaid 1975], is designed to indicate the difficulty of a passage of text according to the following formula:

\[
206.835 - 1.015 \left( \frac{\text{total words}}{\text{total sentences}} \right) - 84.6 \left( \frac{\text{total syllables}}{\text{total words}} \right)
\]

Many readability tests however, are rather basic since they do not consider the semantic difficulty of the words nor do they make any distinction on the language of the words in the text. Thus two open questions for our research are whether or not to include readability scores and if so, which test or formula is most suitable. In her thesis [Eye Tracking to Support eLearning 2016], Copeland presents an extensive list of readability tests that are available, including: COH Metrix [REF], Gunning-Fog [REF], Coleman-Liau [REF], SMOG [REF], and Automated Readability Score [REF]. All of these tests however are still quantitatively computed in a similar fashion to Flesch–Kincaid, except for the COH Metrix test. The COH Metrix is an alternative readability test that computes text coherence, defined as the interaction between linguistic representations and knowledge representations [Grasser 2013].

4.4.3 Properties of the Visual Task

We identified two properties that explain that the operations that the reference’s text establishes on the visualization in the MMD, which can be summarized in terms of two abstract visual tasks as defined in Munzer [2014].

4.4.3.1 Search Task

First, the reference’s text provides the criteria for how the targets in the visualization are to be searched for, based on two subtypes of search tasks, i.e., locate or browse for the targets:

**Locate:** The targets are known since they can be identified directly according to one or more attributes specified by the reference’s text. Almost all locate tasks in the Kong dataset identify targets using a combination of primary key and value, as well as secondary key when applicable (refer to previous Section 4.4.1.1 for how keys are defined). We also identified two additional ways in which a locate tasks can specify targets:
- Locate by visual channel - Targets are elicited according to color or position. E.g.: “the blue bar”, “the leftmost bar”, “the top blue bar”, etc. A visual channel used to control the appearance of marks, such as: position, color (hue, saturation, luminance), shape, tilt, etc.

- Locate by annotation - When the target is elicited by referring to an annotation in the visualization. An annotation refers to the addition of graphical or textual annotations associated with one or more preexisting visualization elements.

**Browse:** The targets are *unknown* prior to searching the visualization since the exact identity of the marks are not all directly specified by the reference’s text. Targets are thus searched for by browsing the visualization according to some criteria specified by the reference’s text, e.g., by filtering the data or describing a trend/pattern in the data.

### 4.4.3.2 Query Task

Second, once the targets have been searched for and found in the visualization, *query* describes the task of extracting the relevant information from the targets. There are 3 query types, defined solely by the number of search targets $t$ in a visualization with $n$ marks:

- **Identify:** One search target: $t = 1$.
- **Compare:** Multiple search targets. Typically more difficult than identify, since it includes identifying multiple targets: $1 < t < n$.
- **Summarize:** The full set of marks in the visualization are targeted: $t = n$.

### 4.4.3.3 Example 1: Locate & Identify Task

Here we provide an example (Figure 17) of a reference that is categorized as a locate and identify task: “76% say that religious extremism is not or not at all common among Protestants in the prisons where they work”. First, the reference’s text specifies three attributes that can be used to locate the target (primary key, secondary key, and value), indicating that the target is known. Next, the target mark is uniquely identified by searching the visualization.

![Figure 17. Example of a single target reference, described by a locate and identify task.](image-url)
4.4.3.4 Example 2: Browse & Compare task

Here we provide an example (Figure 18) of a reference that is categorized as a browse & compare task: “This rose to 84% for those aged 70% and over”. First, the reference describes how to browse for the targets in the visualization, i.e., those aged 70 and over. Notice how the exact identities and number of targets are not known prior to searching the visualization. Once the target marks are searched for in the visualization, they can be compared since there is more than one target.

4.4.3.5 Visual Task Statistics on the Kong Dataset

Here we provide statistics on the references in the Kong dataset according to their visual task properties in Table 5. As the table shows, most locate references have only one target to identify. Also, we found no instances of locate and summarize, likely because it is not practical for one reference to explicitly specify every mark in the visualization since it would be a long list of attributes to process. As for browse, most are compare tasks. We found no instances of browse and identify, likely because it is more suitable for a reference to identify single targets by locate.

<table>
<thead>
<tr>
<th>Search Type</th>
<th>Total</th>
<th>Query Type</th>
<th>Total</th>
<th>Number of Targets:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Avg.</td>
</tr>
<tr>
<td>Locate</td>
<td>171</td>
<td>Identify</td>
<td>144</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Compare</td>
<td>27</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Summarize</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Browse</td>
<td>11</td>
<td>Identify</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Compare</td>
<td>10</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Summarize</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 5. Statistics on the search and query types, as well as the number of targets for the 182 references in the Kong dataset.
4.5 Data Analysis

We plan to perform two stages of analysis on the data collected from the study. First we will perform an analysis pertaining to user experience, followed by an analysis of users’ gaze data.

4.5.1 Analysis of User Experience

This analysis will address the two research questions in Milestone1B:

- **RQ1:** To what extent does providing adaptive highlighting interventions, triggered when the user looks at a reference’s text in the MMD, improve the effectiveness of MMDs?
- **RQ2:** To what extent does the effectiveness of the highlighting interventions in RQ1 depend on the individual differences of the user?

First, we will use a MANOVA [Field 2009] to carry out the analysis for RQ1. The independent measures in the model will contain factors such as intervention properties and reference complexity. We will evaluate all of the dependent measures that we collected from the study, which have not yet been finalized (see Section 4.3). Next for RQ2, in order to uncover any possible interaction effects from user characteristics, a separate MANCOVA [Field 2009] will be applied for each of the user characteristics we measured. This follows an established analysis technique we’ve previously employed for user characteristics in [Carenini 2014] and [Conati 2014].

4.5.2 Analysis of Gaze Data

The purpose of this analysis is to look at user’s eye tracking data to help explain results from the analysis of user experience. In particular, by examining where and how users are looking at the text in the MMD, the visualization, the references, and the interventions, we can potentially explain why significant differences in performance did or didn’t occur. This analysis is still useful in the event that we do not detect any results from the analysis of user experience, because it can still provide us with reasons why (e.g., users were not actually looking at the interventions we provided). We will employ the same methodology that both Toker et al. [2013] and Toker & Conati [2014] used to investigate eye tracking data from two different studies with visualizations. The required methodology consists of running a battery of Mixed Model ANOVAs [Field 2009] with gaze features as the dependent measures. Gaze features are generated from users’ raw eye tracking data, and are used to describe how various elements of the interface/visualization were processed. Due to the large number of Mixed Model ANOVAs, and thus increased likelihood of committing a Type I error, Bonferroni adjustments are then applied to adjust the alpha threshold by grouping similar gaze features into families and adjusting according to each family size (see [Toker et al. 2013] for further details). Results from running all of the Mixed Models, in conjunction with results from the analysis of user experience, may provide insights into how users are visually processing the MMDs and interventions, which would serve to refine and improve how future adaptive support is provided.
5. **MILESTONE 2: Predicting User Characteristics in Real-time**

**RQ3:** *Using eye tracking, to what extent can we reliably predict relevant user characteristics in real-time as the user interacts with MMDs.*

Even though a body of work now exists that establishes the feasibility of making predictions of various user characteristics using eye tracking during information visualization tasks (see Related Work Section 2.3), it has yet to be shown for MMDs, which contain visualization(s) plus a body of text. The main purpose of Milestone 2 is to investigate if eye tracking data can predict relevant user characteristics during interactions with adaptive MMDs. The motivation for this work is to provide a non-intrusive method for predicting user characteristics so that interventions could be personalized to users as they process MMDs. Our plan is to conduct machine learning experiments with the data collected from the study described in the previous Milestone. The tasks for conducting the machine learning experiments are shown in Figure 19. Details regarding the current description and challenges for processing the user characteristics data and training/evaluating the machine learning models are presented in the Introduction (Section 1.3.3). Here, we will present details and challenges for the task of processing eye tracking data.

![Figure 19. Tasks for conducting machine learning experiments in Milestone 2.](image)

### 5.1 Process Eye Tracking Data

This task consists of generating suitable input features for machine learning using the eye tracking data collected from the study in Milestone 1B. Eye tracking data logged from the study will consist of a sequence timestamps indicating the users’ fixations and saccades during their interactions with the MMDs. Fixations are defined as pauses over informative regions of interest and are described by their location (x-y coordinate on the screen) and duration; saccades are defined as rapid movements between fixations and are described by a starting location and end location on the screen [Salvucci & Goldberg 2000]. Next, the fixation and saccade data will be processed with the Eye Tracking and Data Analysis Toolkit (EMDAT: https://github.com/ATUAV/EMDAT) which generates an extensive set of aggregate and summative gaze features to be leveraged for in-depth analysis or machine learning experiments. These gaze features capture various aspects visual processing across the entire interface and include
summative statistics on: the durations of the fixations, the distances of the saccades, the angles between consecutive saccades, and the velocities of the saccades. We provide a list of some of the gaze features that EMDAT generates in Table 6a. In addition, gaze features from EMDAT can also be generated according to Areas of Interest (AOIs). An AOI serves to captures gaze activity within specific bounded geometric regions of the screen. Some of the AOI gaze features that EMDAT generates include: number of fixations in each AOI, number of transitions from each AOI to each other AOI, and longest fixation in each AOI (see Table 6b). Thus, an open question for our research is deciding on how to define AOIs in our MMDs. The challenge in selecting how many AOIs to define is that some AOI features are expressed in terms of every other AOI (e.g., transitions), and thus the number of AOI features grows exponentially as more AOIs are added. Thus, to prevent an overload of gaze features, we are considering defining a minimal starting set of AOIs on the MMD: one AOI over the body of text, one AOI for each reference’s text in the MMD, and one AOI over the visualization (see Figure 20).

Table 6. Example of gaze features generated from EMDAT: a) gaze features describing visual processing according to the whole screen; b) gaze features describing activity within one or more Areas of Interest (AOIs)

Figure 20. Possible starting set of AOIs (Areas of Interest) that we will define over the MMD.
Appendix A. Timeline and Publications

The timeline below (Table 7) shows a high level view of the work carried out from the beginning of the PhD, followed by 24 months of proposed work leading up to submission of the thesis. Descriptions of relevant papers bolded in the timeline will follow.

<table>
<thead>
<tr>
<th>Year</th>
<th>Term</th>
<th>RQ1/RQ2</th>
<th>RQ3</th>
<th>Other Work</th>
<th>Non-PhD Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>Fall</td>
<td>P1: Highlighting Intervention Study, CHI’14</td>
<td>P2: Predicting Skill Acquisition, IUI’14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>Winter</td>
<td>P3: Gaze Analysis of Intervention Study, UMAP’14</td>
<td>P5: Predicting Skill Acquisition with Pupil and Head Distance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td></td>
<td>P4: Predicting Learning Curves, IUI’15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Winter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>Winter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fall</td>
<td></td>
<td></td>
<td>Player Motivation in Games, UMAP’15</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MITACS Internship at East Side Games</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>Winter</td>
<td></td>
<td></td>
<td>Selected Kong dataset</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td></td>
<td></td>
<td>Writing Thesis Proposal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Final class: cs500</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sept</td>
<td></td>
<td></td>
<td>Proposal Defense</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oct</td>
<td>Complete Study Design (includes preliminary pilot studies)</td>
<td>Complete Experimental Platform</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nov</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jan</td>
<td>P6: Publish results on selecting study factors from pilot testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Apr</td>
<td>Conduct User Study</td>
<td></td>
<td>Begin Writing Thesis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>May</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>June</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>July</td>
<td>Analysis of performance. P7: Publish Results (Possible: CHI’18)</td>
<td>Write Thesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sept</td>
<td>Analysis of eye tracking data. P8: Publish Results (Possible: UMAP’18 or AAAI’18)</td>
<td>Write Thesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>Oct</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nov</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dec</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jan</td>
<td>P9: Predicting User Characteristics with Eye Tracking</td>
<td>Write Thesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Feb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Apr</td>
<td></td>
<td></td>
<td></td>
<td>Write Thesis</td>
</tr>
<tr>
<td></td>
<td>May</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>June</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jul</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aug</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Timeline depicting past work and proposed thesis work.
A.1 Existing Publications

Here we present four papers that were published during the PhD (but prior to the proposal), that we believe should count toward the PhD.


This paper designed a set of highlighting interventions that were then tested and shown to be effective at improving user experience (i.e., time on task, perceived usefulness) with bar graph visualization tasks. This paper provides a significant contribution to the work in our proposal since it provides us with an ideal starting set of highlighting interventions that we know were useful with bar graphs. In addition, the paper also showed that the effectiveness of the interventions can depend on the type of task, which provided motivation in our work to define and control for reference complexity. The list of authors for this paper is provided in alphabetical order. My contribution to the paper included: helping design the users study and interventions, running study participants, establishing and running the data analysis, reporting all of the results, and writing part of the paper.


The contribution of P2 is showing that we can use eye tracking predict a user’s skill acquisition state (i.e., a user characteristic) during tasks with bar graph visualizations. This paper provides methodology and support towards achieving a similar objective in RQ3 since our MMDs also contain bar graph visualizations. As first author of the paper, I carried out all of the analysis, reported the results, and wrote the majority of the paper.


The contribution of P3 is showing that we can use eye tracking to analyze a user’s gaze patterns in order to provide possible reasons for why differences in performance existed between users according to their user characteristics. Results in the paper serve to provide guidelines on where in the interface to refine or design new types of adaptive support. For instance, the paper identified that users with low verbal working memory had difficulty processing the text in the task questions, indicating they may need additional support with this part of the interface. P3 provides evidence that we can use the same methodology to provide explanations for the results obtained in RQ1 and RQ2. As first author of the paper, I carried out the analysis, interpreted the results, and wrote the majority of the paper.

The contribution of this paper is showing that it is possible to use eye tracking for predicting a user’s learning curve as they work with an interactive bar graph visualization. Similar to P1, P4 provides methodology and support towards achieving a similar objective of predicting user characteristics for RQ3. As second author of the publication, I assisted in the interpretation of results and identifying and interpreting the top gaze features selected in the models, I wrote sections on these interpretations in the paper accordingly.

A.2 Proposed Publications

Here we present possible publications that could be obtained from completing work in the PhD proposal:

P5. Predicting Skill Acquisition Paper with Head Distance & Pupil Measures

P5 extends the work in P1 by providing details on how to achieve better accuracies for predicting user characteristics from eye tracking data. In particular, P5 identifies a better algorithm (Random Forest), and provides an approach for developing a stricter baseline for comparing predictive accuracy. Additionally, P5 also investigates how other measures collectible from an eye tracker (pupil measures and a user’s head distance from the screen) can be useful for predicting user characteristics. Results show that including these measures does in fact achieve better predictive accuracies. P5 is relevant to our proposal because it provides improved methodology which will be used for RQ3. As first author, P5 has already been written, and we are currently considering suitable venues for submitting it.

P6. Outcomes from preliminary pilot tests to select study factors

P6 would provide an outline of our work on designing user-adaptive MMDs, and in particular would report on results from our pilot tests on selecting factors for the study design (e.g., designing the interventions and crafting the Kong dataset into a suitable form for running our study).

P7. Analysis of User experience

P7 would provide results from running the user study and addresses RQ1 and RQ2, namely, were the interventions effective, and did user characteristics play a role. In the event of a null outcome, we may consider including an analysis of gaze patterns into this paper to provide explanations for this outcome (see Section 1.3.2, Outcome 0 & Outcome 1).

P8. Analysis of Gaze Patterns

P8 would provide results from analyzing user’s gaze patterns from Study 1. The goal would be to provide reasons for the results in P7, and also provide guidelines for improving the interventions as well as designing new ones.
P9. Predicting User Characteristics from gaze data

P9 would report on how well we can predict user differences from gaze data collected from our study with MMDs. In addition, we would report on all of the machine learning experiments conducted to identify how to best make these predictions, and which configurations of machine learning models are best suited (e.g., which gaze features, which algorithms, etc.). If successful, the outcome would provide grounds for future work, namely, the means for a user-adaptive MMD system to predict individual differences in real-time in order to guide the selection of the interventions accordingly.
Bibliography


