Detection of Recognition Errors from Eye Gaze Behavior in Sketch Recognition Interfaces

by

Özem Kalay

A Thesis Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science
in
Computer Engineering

Koç University

August 2014

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Özem Kalay

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

____

Assist. Prof. T. Metin Sezgin (Koç University)

____

Assoc. Prof. Engin Erzin (Koç University)

____

Assist. Prof. Albert Ali Salah (Boğaziçi University)

Date: ____________________

To my grandmother

ABSTRACT

Sketch based intelligent interfaces are gaining popularity as pen based hardware becomes more widespread. These interfaces make use of sketch recognition technology to facilitate natural and efficient interaction. Nevertheless all sketch recognition systems suffer from misrecognitions, which inflicts a correction cost on to the user. Every time a symbol gets misrecognized, the user explicitly or implicitly signals his intention to correct the error, and does so by either redrawing the symbol or selecting it from a list of alternatives. We propose a system for alleviating the cost of this two-step process by detecting users' intention to fix misrecognitions based on their eye gaze activity. In particular, we show that users' natural reaction to misrecognitions manifests itself in the form of characteristic eye gaze movement patterns. Furthermore, these patterns can be used to read users' intention to fix errors before they initiate such action. We have three main contributions. First, we present a carefully constructed Wizard of Oz setup for recording eye gaze patterns under two sketch-based interaction conditions. Then, we present a set of gaze-based features, which were designed to capture qualitative characteristics of users' eye gaze behavior. Finally, we present a framework for recognizing users' intention to fix errors, which achieves an 86% prediction accuracy. We support our findings through detailed experiments and statistical analyses, which provide further insight into how much can be inferred from eye gaze patterns that naturally emerge during pen-based interaction.

ÖZET

Kalem temelli donanımlar yaygınlaştıkça, çizim temelli akıllı arayüzler de daha popüler hale gelmektedir. Bu arayüzler çizim tanıma teknolojisinden faydalanarak doğal ve verimli insan-bilgisayar etkileşimi sağlamaktadır. Fakat bütün çizim tanıma sistemleri aynı zamanda çizim tanıma hataları da yapmaktadır ve bu hataları kullanıcılar düzeltmek durumundadırlar. Bu durum kullanıcının üzerine bir düzeltme yükü bindirmektedir. Bir çizim yanlış tanındığı zaman, kullanıcı hatayı düzeltme niyetini açıkça ya da üstü kapalı bir şekilde belli etmektedir. Ardından ya çizimini yeniden yaparak ya da bir sahneye yerleştirmek istediği sembolü bir sembol listesinden seçerek hatayı düzeltmektedir. Bu çalışmada çizim tanıma hatalarını düzeltmenin yükünü hafifletmek için, kullanıcının çizim tanıma hatasını düzeltme niyetini bakış verisinden faydalanarak belirleyebilen bir sistem sunuyoruz. Kullanıcıların çizimlerin yanlış tanınması halinde gösterdikleri tepkilerin karakteristik göz hareketlerine yol açtığını gösteriyoruz. Ayrıca bu göz hareketlerinin kullanıcılar henüz düzeltmeye başlamadan önce onların düzeltme niyetlerini okumak için kullanabileceğini gösteriyoruz. Bu çalışma 3 temel katkı sağlamaktadır. İlk olarak iki farklı çizim temelli etkileşim senaryosunda göz hareketlerinin kaydedildiği ve özenle kurgulanmış bir "Wizard of Oz" (Oz büyücüsü) deney düzeneği sunuyoruz. Daha sonra kullanıcıların göz hareketlerinin nitel karakteristiklerini ifade eden bir öznitelik seti sunuyoruz. Son olarak kullanıcıların çizim tanıma hatalarını düzeltme niyetlerini %86 doğrulukla tespit edebilen bir makine öğrenmesi taslağı sunuyoruz. Bulgularımızı kalem temelli etkileşim sırasında oluşan doğal göz hareketlerinden ne kadar anlam çıkarılabileceğini gösteren ayrıntılı deneyler ve istatistik analizler ile destekliyoruz.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Asst. Prof. Tevfik Metin Sezgin for his invaluable guidance, support and patience throughout my studies. He was always available when I had questions and I feel lucky to have benefited from his scientific mentorship, knowledge and ideas. Working with him taught me how to approach scientific questions as well as how to effectively present my findings.

I am indebted to Asst. Prof. Albert Ali Salah and Assoc. Prof. Engin Erzin for agreeing to be in my thesis committee and contributing with their suggestions and comments.

I am forever grateful to my supporting and understanding family; my mother for being a great pillar of support, my father for sharing his serenity and wisdom, my sister and Jonathan Gruber for their valuable advice and help during my research as well as for being my confidants and my brother for always loving and believing in me.

I would like to give special thanks to Uğur Aygün for being with me through thick and thin and for letting me lean on him when I needed it the most. His advice and enthusiasm helped me do my research with greater motivation.

I also owe many thanks to my officemates and members of IUI lab; Neşe Alyüz, Banuçiçek Gürcüoğlu, Burak Özen, Çağla Çığ, Kemal Tuğrul Yeşilbek, Şerike Çakmak, Cansu Şen, Atakan Arasan, Erelcan Yanık, Burak Özaydın, Deniz Can Yıldırım, Bekir Berker Türker and Shabbir Marzban who had to listen to my concerns almost daily and gave me solace and support. Also I would like to thank Ayşe Küçükyılmaz, Sinan Tümen and Çağlar Tırkaz for their advice and support.

This thesis is funded by TUBITAK 110E175 project.

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES x

NOMENCLATURE xii

Chapter 1 INTRODUCTION 1

Chapter 2 RELATED WORK 8

Chapter 3 DATA COLLECTION 14

3.1 Design Choices 14

3.2 Physical Set-up 16

3.3 Task 19

3.3.1 Creating Contextually Inferable Misrecognitions 20

3.3.2 Creating Contextually Non-conflicting Misrecognitions 21

3.4 Procedure 21

Chapter 4 FACTORS ON CORRECTION PERFORMANCE 23

4.1 Effect of Sketch Misrecognition's Contextual Inferrability 23

4.2 Effect of the Sketch Misrecognition's Visual Field 25

Chapter 5 FEATURES 28

5.1 Gaze Behaviors 28

5.2 Preprocessing 30

5.3 Features 30

5.4 Distribution of Features 33

Chapter 6 CLASSIFICATION 35
6.1 Classification Framework 35
6.2 Effect of Background Information on Classification 37
6.3 Effect of Visual Area of Stimuli on Classification 38
6.4 Prediction as a Function of Time 39
6.5 Feature Contribution 41
Chapter 7 CONCLUSION AND FUTURE WORK 47
APPENDIX 49
BIBLIOGRAPHY 71

LIST OF TABLES

Table 5-1: Features 31

Table 6-1: True positive, true negative, false positive and false negative rates 36

Table 6-2: Accuracy results for CI and CN data sets 37

Table 6-3: Feature groups 41

Table 6-4: Feature group combinations. ‘0’ indicates that the feature group is absent, while ‘1’ indicates that the feature group exists in the combination 42

LIST OF FIGURES

Figure 1-1: Life cycle of a sketch in a sketch-based system 2
Figure 1-2: Yarbus's study 3
Figure 3-1: The positions of the participant and the wizard 16
Figure 3-2: The participant interface 17
Figure 3-3: The wizard interface 18
Figure 3-4: An example segment of the progress of clock filling task for the first part of the experiment. 20
Figure 3-5: An example segment of the progress of clock filling task for the first part of the experiment 20
Figure 4-1: The distribution of noticed and missed sketch misrecognitions by the relation of the misrecognition with the context (inferable of non-conflicting). 24
Figure 4-2: Visual areas of the human visual field 25
Figure 4-3: Correction rates of sketch misrecognitions in different visual fields when they are contextually inferable 26
Figure 4-4: Correction rates of sketch misrecognitions in different visual fields when they are contextually non-conflicting 26
Figure 5-1: Example gaze paths for (a) Direct Comeback (b) Double take and (c) Investigation behaviors 29
Figure 6-1: Change of correct classification rate by the visual area 39
Figure 6-2: Change of prediction accuracy by prediction time prior to correction........ 40
Figure 6-3: Marginal mean changes for each feature group 43
Figure 6-4: Effect of interaction between Duration and Frequency feature groups on classification accuracy 44
Figure 6-5: Effect of interaction between Duration and Distance feature groups on classification accuracy 45

Figure 6-6: Effect of interaction between Frequency and Distance feature groups on classification accuracy.. 46

NOMENCLATURE

SVM	Support Vector Machines
AOI	Areas of Interest
HMM	Hidden Markov Models
DBN	Dynamic Bayesian Networks
DMM	Dirichlet Mixture Models
CI	Contextually Inferable
CN	Contextually Non-conflicting

Chapter 1

INTRODUCTION

Sketch recognition is the segmentation and interpretation of sketches by the computer in a human-like manner. Unlike plain pen-based interfaces, which see pen input merely as a collection of ink coordinates, sketch-based intelligent interfaces are able to interpret hand-drawn sketches. A sketch-based intelligent interface segments user sketches into meaningful pieces and labels each segment as an object. For example, in a sketch-based intelligent circuit simulator, the interface receives electrical circuit diagrams as input, segments the sketch into circuit elements and labels each segment. The recognition results can then be fed to an off the shelf circuit simulator to display circuit behavior. The interface provides the users with recognition feedback on the label of each sketch segment. Sketch-based interfaces employ a variety of feedback strategies to indicate recognition results. The most common strategy is replacing sketch fragments with a recognized version of the intended object. Sketch-based intelligent systems provide natural and efficient interaction and they are becoming more popular as pen based systems get more widespread. These systems have started appearing in applications developed for educational [1], [2] and design purposes [3]. However, even the most advanced sketch recognition systems suffer from sketch misrecognitions. A sketch misrecognition is the incorrect labeling of a sketch segment, such as labeling the capacitor element on the circuit diagram sketch as a battery. In current sketch recognition systems the user generally handles these misrecognitions by erasing the misrecognition feedback and remaking the sketch or taking any other step to acquire correct sketch recognition feedback. These extra steps waste time and impact efficiency.

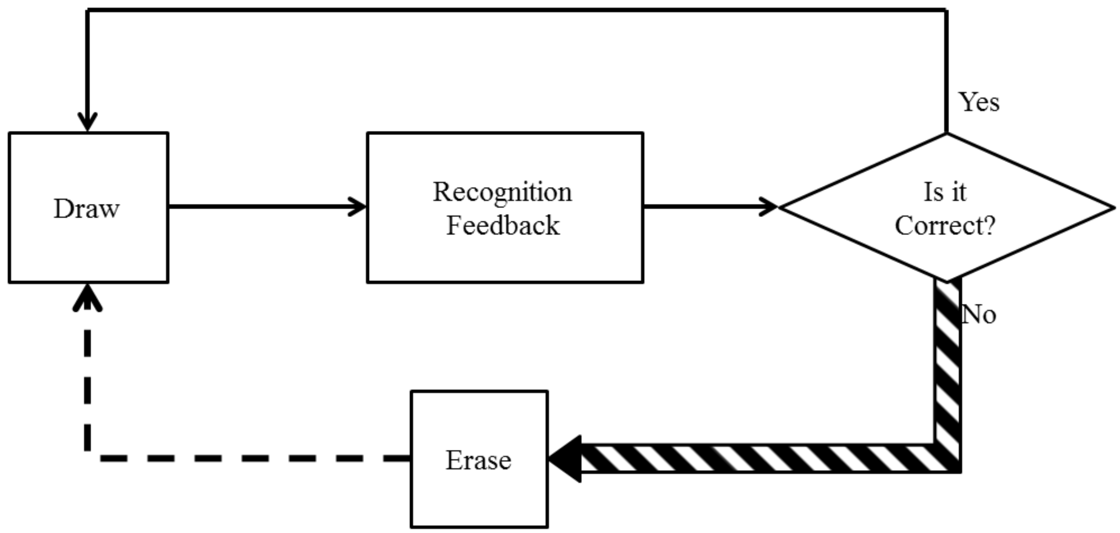


Figure 1-1: Life cycle of a sketch in a sketch-based system

A system that aims to offer help with the correction of a misrecognition must first detect misrecognitions. State of the art systems detect sketch misrecognitions only after the user takes action to correct them, which is often too late. In this thesis, we propose a system that alleviates the cost of recovering from misrecognitions by detecting the occurrence of sketch misrecognitions before user takes action.

In order to further explain the problem and our motivation we give life cycle of a sketch in Figure 1-1. Cycle starts with the users' drawing and continues with recognition module's feedback. If the feedback is correct the cycle goes back to the beginning as the users continue their drawing. However if the feedback is false users first have to erase the sketch and then remake it at which point the cycle goes back to the beginning for the same sketch. Our aim is to detect that the sketch recognition feedback is false between the feedback's occurrence and the time users erase it. This process is shown by the arrow with diagonal stripes. By detecting the false feedback early on, we will be able to give users sketch misrecognition fixing support before users themselves start to fix it. This will allow us to get rid of the excess part of the cycle which is denoted by the dotted line.

Sketch-based interaction has two parties; the user and the sketch-based system. Since sketch-based system is the source of the misrecognition problem it cannot also constitute the solution. Hence we look for the solution at the other party; the user. We make use of users' reactions towards sketch recognition feedback to detect the occurrence of misrecognition. Specifically we employ the gaze modality for this purpose.

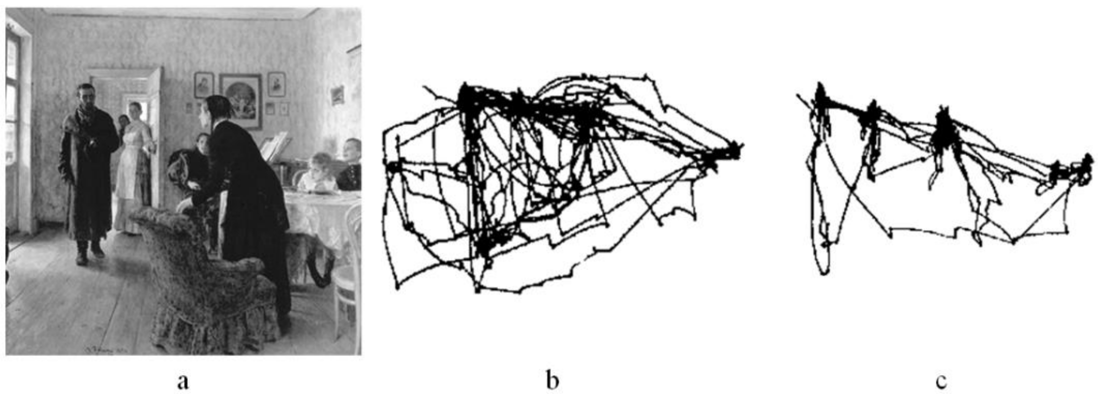


Figure 1-2: Yarbus's study (a) The painting shown to participants (b) An example to the gaze trajectories when participants viewed the painting freely (c) An example to the gaze trajectories when participants tried to assess the ages of people in the painting [4]

One of the first studies on human eye gaze belongs to Yarbus. In his study [4] Yarbus showed the painting seen in Figure 1-2-a to the participants and asked a number of questions about the painting. When participants were asked to view the painting freely their gaze displayed a trajectory similar to the one seen in Figure 1-2-b. When they were asked to estimate the ages of people in the painting the gaze trajectories were similar to the one in Figure 1-2-c. This study showed that people's eye gaze is shaped according to their goals. Yarbus used an invasive eye tracking system. Fortunately, owing to developing technology eye tracking systems became non-invasive and low cost. This fueled the research on eye gaze and enriched research in psychology and computer science domains further. Eye gaze is believed to be a good indicator of people's attention [5] and their future intentions with its proactive behaviors [6].

In this study, we take advantage of the richness of the gaze modality to predict users' intention to fix sketch misrecognition before they take action. We hypothesize that users display distinctive gaze behavior when they realize a misrecognition and move on to fix it. By attending to the characteristics of the eye gaze during this time, the interface can infer whether the sketch recognition has failed or succeeded.

Research [7] show that the deployment of attention and therefore gaze can be modeled by a two party framework consisting of top-down and bottom-up processes. *Top-down* processes are produced by the task at hand and memory while *bottom-up* processes are produced by visual properties of objects on the scene such as color or contrast. Attentional models focused on bottom-up processes [7] for a long time, independent of top-down processes. Only recently top-down processes started to be included in attentional models [8], [9]. In this work our aim is to observe how the task of fixing sketch misrecognitions affects gaze behavior, minimizing the effects of bottom-up processes. Hence we designed our study to contain simple shapes with minimum visual saliency in order to capture unbiased top-down behaviors.

Considering the fast development of eye tracking technology, we envision that eye trackers will be common built-in accessories in mobile devices. We believe that, for the sake of comfort during the interaction, technology will avoid forcing the users to use any additional hardware. In accordance with this vision we conducted our study with a desk mounted eye tracker, simulating the future gaze-based interaction.

We also investigate which factors may affect users' gaze behavior as they notice and move on to fix sketch misrecognitions. Sketch misrecognitions can effectively be thought of as local changes made in user's drawing. Hence, a body of research that can light the way to the factors that affect sketch misrecognition detection is the work on change detection. Fueled by the discovery of change blindness phenomenon [10], change detection research focuses on humans' performance to detect a change made to a scene. This body of research aims to gain insight on the way humans construct visual representations of their environment in their minds by investigating the effects of different factors on change detection performance. It is proved that a variety of factors affect change detection performance in humans such as the onlookers' attention level [11], familiarity [12][13] or expertise [14] on the context of the scene. Also angular distance of the change according to gaze [15][16], change's semantic consistency with the scene [17] or the type of change [18] (if the change is induced by appearance of a new object or disappearance of an existing object) affect how successfully can people detect changes. We hypothesize that factors affecting change detection can also be influential in sketch misrecognition detection and gaze behavior during this process.

Research shows that detection of changes is associated with the relation of the change with the context [12][13][14][17]. Hence drawing inspiration from this line of work we extended our research to contain the factors of whether the sketch misrecognition is *contextually inferable* or *contextually non-conflicting*. Misrecognitions become *contextually inferable* if they disturb the contextual coherence of the scene. In this case people can conclude that the sketch recognition is false suspecting that there should have been another object in the place of the misrecognized one if the recognition was correct. This effect can increase as the misrecognized object's dissonance with the scene increases. It can also increase as the onlookers' information on the object domain increases since it makes differentiating contextual incoherence easier. On the contrary, *contextually non-conflicting* misrecognitions, as their name suggests, do not disturb the contextual coherence. We investigated how users respond to these two different types of misrecognitions through their eye gaze.

We also investigated the effect of angular distance between the gaze and the sketch misrecognition at the time of occurrence of sketch misrecognition. During natural interaction the sketch misrecognitions can occur at different angular distances from the user's gaze. In this work we investigated the effects of angular distance under the context of human visual fields. Humans' visual field is divided into 3 main areas; foveal, parafoveal and peripheral. Foveal area is the center of the visual area, while peripheral area is the outermost part. Parafoveal area is a transition field between the foveal and peripheral areas. Eccentricity increases as we go from foveal to peripheral area. Each visual field has similar visual quality in itself; people can pick up more color and detail in the foveal area while they notice moving objects in their peripheral area more easily.

It is important to note that unlike sketch scenes, the scenes used in change detection research are ones that the people has no interaction or effect on. A sketch scene is constructed by people, while scenes used in change detection research have been limited to pre-made images. In addition, our aim is not only to measure sketch misrecognition detection performance but to observe the gaze behavior prior to detection. There are no studies on user's gaze behavior as they intend to fix a sketch misrecognition. Hence we designed an experiment that simulates the natural interaction. We built a sketch recognition simulator that creates sketch misrecognitions and records eye gaze during this interaction. We included factors of sketch misrecognition's contextual inferrability and visual field into the experiment by creating adequate conditions.

Chapter 2 provides necessary background and literature review on user interfaces utilizing gaze modality. It contains works on intelligent systems that utilizes gaze to understand users' mental state, needs or intentions. Chapter 3 describes user experiment and in particular user task, procedure and the experiment setup. It gives explanations for design choices that shaped the experiment. Analyses on the effect of contextual inferrability and angular distance on user's performance of realizing sketch misrecognitions are presented in Chapter 4. Chapter 5 contains our observations on the collected gaze data that give information on the participants' behavior as they intend to correct sketch misrecognitions. These observations provide insight into the data. Thereafter we present the feature set and explain how we constructed this set. We give the intuition for each feature. Chapter 6 gives detailed description of machine learning framework built for discriminating correction intending gaze behavior. We explain how contextual inferrability and angular distance of sketch misrecognitions affect classification accuracy. The thesis is concluded with a short summary of the performed study and future research work.

Chapter 2

RELATED WORK

Gaze based human-computer interaction has been an active research area for more than 20 years. In this chapter we summarize the work in this area and give details on the studies related to our work.

We find it useful to introduce the terms *fixation* and *saccade* before going into detail since these terms are commonly used in the majority of the studies. A fixation is pausing of gaze at a certain location and saccade is a rapid eye movement between two fixations.

Gaze based interaction systems are divided into two main categories; command and non-command interfaces. Command interfaces use eye gaze as a computer input not different from a mouse pointer. Researchers have built systems that used gaze for object selection, object moving, text scrolling [19] and typing [20]. However this kind of interaction brought the problem of “Midas Touch”. The interface performed actions wherever the user looked at even though the user did not intend for an action. Even though eye gaze was fast and immune to fatigue, it proved to be inappropriate as a means of input for tasks requiring accuracy [21]. Recently researchers have started designing unnatural gaze gestures that will not be confused with natural eye movements to overcome the “Midas Touch” problem and increase the accuracy of gaze as a pointer input [22]–[24]. However this approach is undesirable because it interrupts the gaze’s natural task of observing and gathering and causes high cognitive load [25].

On the other hand, non-command interfaces, including our work, do not force users to manipulate their gaze in an unnatural manner, but observe gaze in its natural flow. Non-command interfaces are generally based on cognitive aspects of eye gaze behavior rather than the motor abilities of the eyes. This line of work coincides with our approach for inferring users' intention of correcting sketch misrecognition. Hence we give a more detailed summary on the studies in this field.

One of the active research areas on natural gaze-based interaction is used for modeling and recognizing the user's activity. Activity recognition research is divided into two main branches by the activity being in real-world or virtual-world (computer screen). Real-world activities generally consist of daily tasks which require hand-eye coordination. In a recent study Fathi et al. [26] introduced a system that classified 25 actions performed during the preparation of different food recipes. They used features based on object information, fixation locations on the object and information on future manipulation of the object. They acquired 47% classification accuracy with Support Vector Machines (SVM). This line of work is useful for our study because it shows how to collect natural gaze data and process it. However we are interested in users' interactions with user interfaces in the virtual world.

Virtual activities consist of tasks that are performed through a computer. Users have minimal physical activity limited only to moving fingers and eyes. Campbell et al. [27] conducted one of the first studies using a virtual task; reading on a computer screen. They created a model that used gaze directions to distinguish reading and searching activities. Bulling et al. [28] conducted 3 different experiments to monitor gaze during 3 different groups of activities. These were activities that take place in an office (copying, reading, writing, watching a video and browsing the Web), reading task (reading and other gaze movements on the page) and visual memory recall (familiar and unfamiliar faces). They showed that gaze patterns are distinctive for subtasks in each group and acquired satisfying classification accuracies with SVMs. Another notable work is by Courtemanche et al. [29] who attributed gaze patterns to areas of interest (AOI). They combined gaze with keyboard and mouse inputs on the screen to predict users' task. Users performed 3 tasks on Google Analytics; evaluating trends in a certain week, new visits and overall traffic. The system acquired 51.3% average accuracy with Hidden Markov Models (HMM). Steichen et al. [30] created a similar system that predicted users' task while working on bar and radar graphs. They also used gaze patterns attributed to AOIs but also used fixation based features such as fixation duration and count. A logistic regression model classified five information visualization tasks with 63.32% average accuracy. These studies are notable for correlating gaze with a task and then interpreting the gaze data. However they all use gaze patterns occurring while the users perform an action. In our study the challenge is to predict users' intention before they start the action.

There are 2 notable studies on prediction of users' intention as decide whether to interact with the user interface. Bader et al. [25] built a probabilistic model based on fixation location respective to the single virtual object on the scene. Their system predicted whether the user intends to select the virtual object or not with 80.7% average accuracy. Despite this high accuracy the interface is quite simple with only one object on the scene compared to our work. Bednarik et al. [31] conducted an experiment where users played a gaze based tile game. They used fixation properties and pupillary responses with SVM to predict whether the user intends to issue a command (move a tile) or not with 76% average accuracy. This work constitutes an example for our work, though it has low accuracy.

We want to detect users' intention to correct sketch misrecognitions. Hence we examined the work on attentive user interfaces which draws motivation from inferring users' interests and demands. Starker and Bolt [32] developed one of the first examples to these interfaces using eye gaze. The interface inferred the user's interest on items which are presented on a virtual planet. If an item (such as a staircase) was determined to be of interest to the user through the user's eye gaze then a simple facial agent commented on the item which was of interest to the user. Hyrskykari et al. [21] described a system that can help users when they encountered a problem during a translation task although they did not implement it. Maglio et al. [33] introduced a system that observes eye movements while the user views web pages and determines if the user is browsing or reading. If the user is reading then the web page is decided to be relevant to the user's interest. The systems continued its search on this topic and present the results to the user. Qvarfordt and Zhai [34] developed an interactive tourist information system. They first collected gaze data with a "Wizard of Oz" study where users interacted with simulated tourist information interface. Then they built a system which gives information about the landmarks that the user is interested in. The system deduced that the user is interested in a landmark by thresholding a parameter combined of properties such as; gaze duration, frequency of gaze reentry and information on previous interest to this landmark. The system was evaluated with a user study and received good comments generally. These works were the pioneers of attentive user interfaces and generally focused on inferring the users' interest. On the other hand our aim is to predict the user's decision about a sketch recognition feedback by discriminating different types of interest; interest shown to misrecognitions and correct recognition feedbacks. This aspect of our work makes it original. These studies also investigated eye gaze in static scenes. However we examine gaze while users interact with a scene which they build themselves and also dynamically changes.

A number of studies investigated how users' interest is manifested in eye gaze in dynamically changing scenes. Hirayama et al. [35] presented a model for determining the objects that are most interesting to the users in a dynamically changing interface. The time passed looking at an object after a new object appeared on the interface was shown to be the most informative cue of the user attention. The authors Prendinger et al. [36] presented a system that detects users' interest to individual objects in a dynamically changing presentation. They built a virtual showroom in which 3D virtual agents presented consumer products and shaped the presentation according to users' interest. The system used Dynamic Bayesian Networks (DBN) that take gaze dwell times and contextual relevance of objects as input infer user users' interest in the presentation. Nakano and Ishii [37] built a system that used eye gaze patterns to determine whether the user is engaged in the conversation with a virtual agent. If the user was disengaged, the agent gave responses that attract users' attention back to conversation. A "Wizard of Oz" experiment was conducted to monitor how users' gaze fell on related or unrelated locations to the conversation. Using these location patterns researchers built a system that autonomously senses disengagement and gives responses to the user. The system was evaluated qualitatively in a second experiment and received generally positive responses. These works are important in the sense that they infer users' interest in dynamically changing scenes. We also investigate users' eye gaze in dynamically changing scenes but in our case the user also play an active role in changing the scene by sketching. Also as we mentioned earlier our aim is to discriminate different types of interest shown to correct and false sketch recognition feedbacks instead of assessing the general interest level of the user. We also have to make this discrimination before the users fix the misrecognition themselves creating a time restriction.

Because gaze intensifies at locations that people are interested in, researchers investigated the possibility of using gaze for inferring the relevance of search items in information retrieval. Researchers conducted user experiments for different information retrieval scenarios and developed eye gaze models for inferring objects' relevance to user's interest. Puolamaki et al. developed and compared different models that predicted relevance of given article titles from users eye gaze. [38] They extracted features based on fixations to each word in article titles, their duration and reading behavior and trained HMMs for irrelevant and relevant titles. They acquired classification accuracy as high as 85.7% by using a Dirichlet Mixture Model (DMM) that combined HMMs and a Gibbs URP (user rating profile) model that represents users' attitudes. The authors also conducted a study on image retrieval [39]. They collected gaze data while users looked for the sports related pages among a number of given pages. Using linear discriminant analysis with features such as fixation duration and count as well as number of fixation transitions between images; they acquired 84.3% precision and 61.4% recall for determining the relevant pages. Papadopoulos et al. [40] presented region based image retrieval which estimated the objects or object parts in each image that the users are interested in. They introduced two new feature extraction techniques; fixation distribution on image parts and coefficients of Discrete Cosine Transform on fixation sequence for acquiring respectively the spatial and temporal properties of the gaze. Using Support Vector Machines (SVM), their system classified objects that were related and unrelated to users' interest with the accuracy of 69.71%. Gaze based information retrieval systems such as these are related to our work through their attempts to model user interest in a given scene. On the other hand in our case users are in a more complicated interaction scheme than just observing a static scene as in information retrieval.

Chapter 3

DATA COLLECTION

In order to assess how the eye gaze behavior changes in reaction to sketch misrecognitions we constructed a carefully designed data collection setup. Our setup involves a sketch-based interaction scenario where the users are subjected to sketch misrecognitions. In this chapter we first discuss factors that we took into consideration while designing our data collection setup. Then we describe setup and user instructions in detail.

3.1 Design Choices

Our aim is to capture natural eye gaze during sketch-based interaction. The most straightforward way to capture natural gaze behavior during sketch-based interaction is to record users' eye gaze while interacting with a real sketch-based intelligent user interface. However in a real sketch recognition system, occurrence of misrecognitions is dependent on the users and the quality of their sketches. On the other hand in our case, we would like to control the conditions under which misrecognitions occur in order to study how these conditions affect user behavior. Hence we built a setup where we could control the location and rate of misrecognitions. For this reason, we used an interface without a real sketch recognition module. The interface, instead, gave sketch recognition feedbacks according to a scenario in which we defined locations and types of recognition feedbacks for all sketches beforehand.

As long as the participant made the sketch correctly to assigned place, the interface was able to give the correct recognition feedback and create the impression of a real sketch recognition system. However participants were prone to mistakes such as making a different sketch from the announced one or sketching to a different place. Also there were still decisions to be given by a human such as when the sketch scene was completed and when to proceed to the next step through the experiment. For handling these issues, we used the Wizard of Oz technique. The user interface was connected to another "wizard" interface which was on a separate computer. During the experiment a wizard observed, and, if necessary, controlled the participant's sketch recognition interface through wizard interface. This way the user interface gave 'intelligent' and consistent responses to all user actions. Figure 3-3 shows the interface used by the "wizard".

The interface intentionally gave misrecognition feedbacks for some of the sketches. Following the work of Wais et al. [41] we arranged the ratio of misrecognitions to be %17 just like in a real sketch recognition system. The interface gave sketch recognition feedback (stimulus) by beautifying the sketch. We designed the misrecognition feedback for each object to be consistent e.g. "3" was always misrecognized as "8" each time it was misrecognized for the interface to be realistic [41].

The experiment consisted of two parts. In the first part we observed people's intention of correcting sketch misrecognition gaze behavior when the misrecognitions were contextually *inferable*. In the second part we created contextually *non-conflicting* sketch misrecognitions and collected gaze data. In the following section we explain the tasks in more detail.

We also controlled the distance between the user's gaze and stimulus at the time of stimulus occurrence. The interface controlled the gaze - unrecognized sketch distance and gave the stimulus after the distance had the desired value. We adjusted consecutive sketch locations to be at a predetermined distance from one another. This way it became highly probable that the distance between gaze and unrecognized sketch would take to the desired value. We used two different sizes of clock templates to cover a wider range of distances.

This way we acquired correction intention and no correction intention gaze behavior for different gaze–stimulus distances.

3.2 Physical Set-up

Figure 3-1 shows the physical set up of the experiment. The set up consisted of 2 computers, one for the participants and one for the "wizard". The two are placed on desks that are perpendicular to each other.

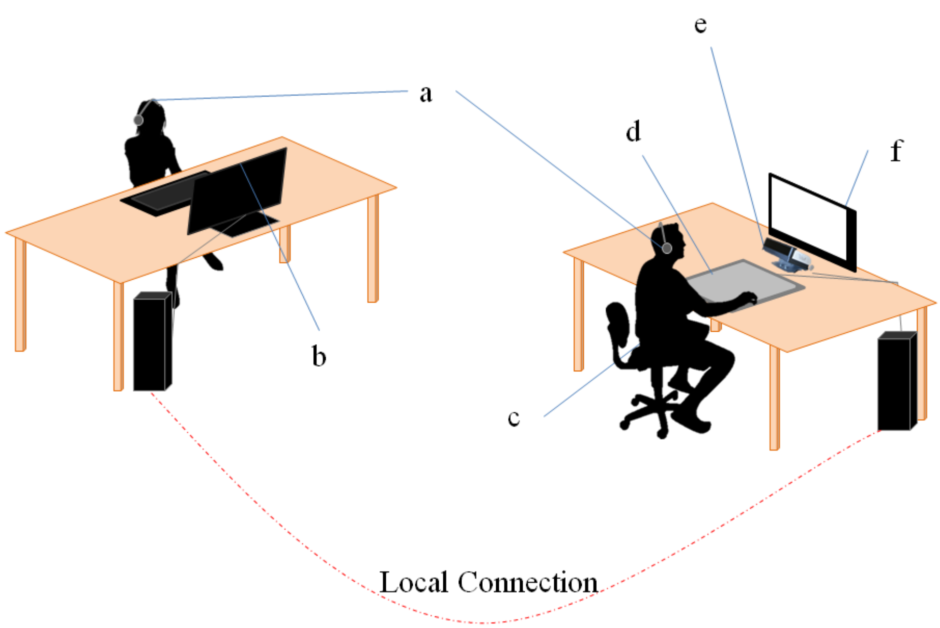


Figure 3-1: The positions of the participant and the wizard (a) Headphones (b) Wizard of Oz Control System (c) User (d) Pen Surface (e) Eye Tracker (f) User Interface

User set-up consisted of a digitizer tablet, a Tobii X120 stand-alone eye tracker and a 19" desktop screen connected to a computer with Intel® Core™ i7, 2.93 GHz processor and 4GB RAM. Wizard set up consisted of a standard keyboard, mouse and 19" screen connected to a computer with Intel® Core™ i7-2700K 3.5GHz processor and 8GB RAM.

Experiment interface established a local connection between user and wizard set-ups. We obtained the experiment software by adding a gaze tracking module on SketchWizard [42].

The users made their sketches on the digitizer tablet but saw the outcome on the screen on the table. The reason for the setup to be this way was that the eye tracker was designed to be placed under the screen on which the gaze is tracked. If we used this screen also for sketching, participants' hand would block the eye tracker. Hence we had to use a separate pen surface and place the eye tracker between the surface and the screen.

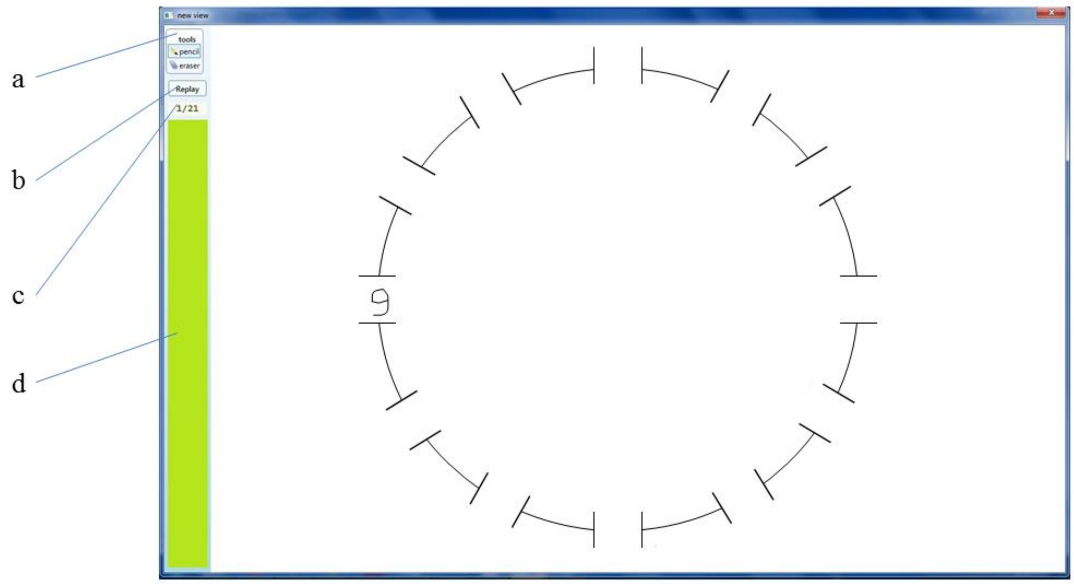


Figure 3-2: The participant interface (a) Tools for sketching and erasing (b) Button for listening the last announcement again (c) Gauge showing number of completed clock templates (d) Color bar showing the validity of gaze data

The participant's interface is in Figure 3-2. We used a vague color for the template so that it would not affect scene perception and users could focus on the sketch recognition feedbacks e g. numbers and shapes.

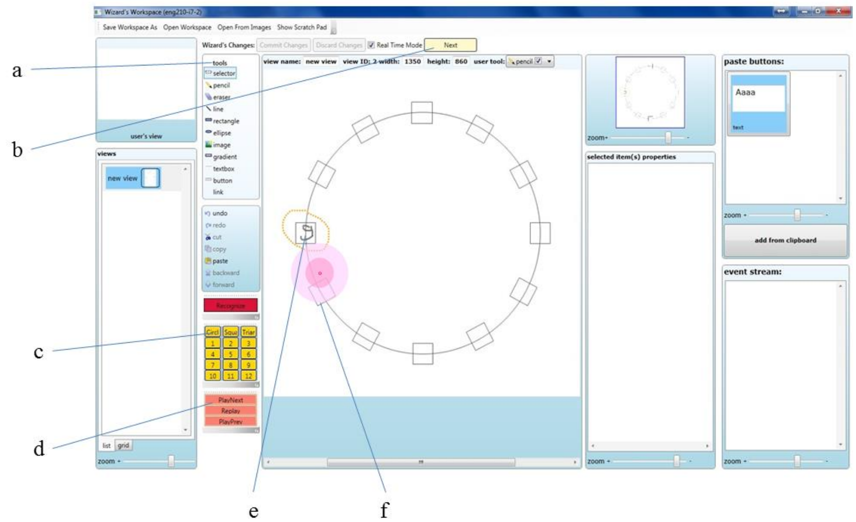


Figure 3-3: The wizard interface (a) Tools for sketching, erasing, selecting and adding new objects to user's interface (b) Button for passing to the next template (c) Individual recognition buttons for each object (d) Buttons for manipulating the order of announcements (e) User's selected sketch (f) User's gaze in real time

The participants were able to erase their strokes although they were unable to erase sketch recognition feedbacks. They could repeat the announcement if they heard but could not understand the command. A gauge showed the number of clock templates completed to inform the participants of their progress. There was also a color bar that appeared green when the eye tracker was able to collect valid data and red when the data was undependable. The data became undependable when the users positioned in a way that the eye tracker could not see the users' eyes. Because participants did not realize the color change of the bar most of the time while they were immersed in their task, a probing sound was added to the system. The participants heard this sound if their gaze data stayed invalid for more than 1 second. Then the interface gave the warning sound to probe the user to adjust their position.

The wizard's interface is seen in Figure 3-3. The wizard watched the participants' actions and gaze in real time. The wizard interface offered the wizard a wide variety of tools. Using these, the wizard was able to select any group of strokes and give sketch recognition for the selected group. The sketch recognition feedbacks occurred only after the gaze-sketch distance reached the intended value. Wizard interface contained a button for each object in the domain. These buttons were used when the participants made the sketch to the wrong place or the wrong object. The wizard gave feedbacks for these sketches unless the user realized and erased the sketch. The wizard also decided when to move on to the next empty template. When the current clock template was full and all the sketches were recognized, the wizard moved to the next template. Both the wizard’s and the participants’ interfaces were recorded during the experiment.

3.3 Task

The tasks in both parts consisted of filling a given wall clock template with instructed objects from headphones though sketch domains were different. Different domains helped us to create contextually inferable and non-conflicting misrecognitions. Through both parts the experiment interface had a clock template on the background as shown in Figure 3-2, though paler in color. The color of the template was made darker in the figures in this thesis to make the viewing easier. With the template we were able to describe the location for each object. Through headphones participants were given audio instructions indicating what object to draw, and where to draw it. The order of the instructions was the same over all participants through entire experiment. Both sketch domains consisted of simple shapes that were easy to sketch and with minimum effect on bottom-up processes.

We requested the participants to make a mark by drawing a line over the misrecognition when they noticed it. Thus, the participants both confirmed that they noticed the misrecognition and marked the time that they would make an action for correcting it. Because of this, we refer the time that participants started to make a pen mark as the beginning of the correction attempt.

3.3.1 Creating Contextually Inferable Misrecognitions

The participants were instructed to draw numbers from 1 through 12 on the clock template in a predetermined mixed order. Participants sketched the instructed number to its place on the template. As the participant made the first stroke of the sketch, instructions for

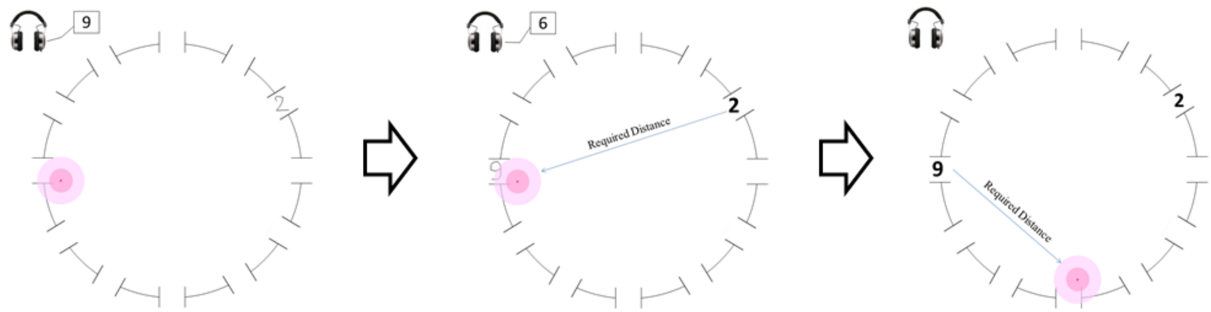


Figure 3-4: An example segment of the progress of clock filling task for the first part of the experiment. The pink circle shows the users eye gaze. Note that the following announcement is made as the user sketches the current sketch. Also note that the sketch recognition feedback is given when the gaze moved to the place of the following sketch satisfying the required distance.

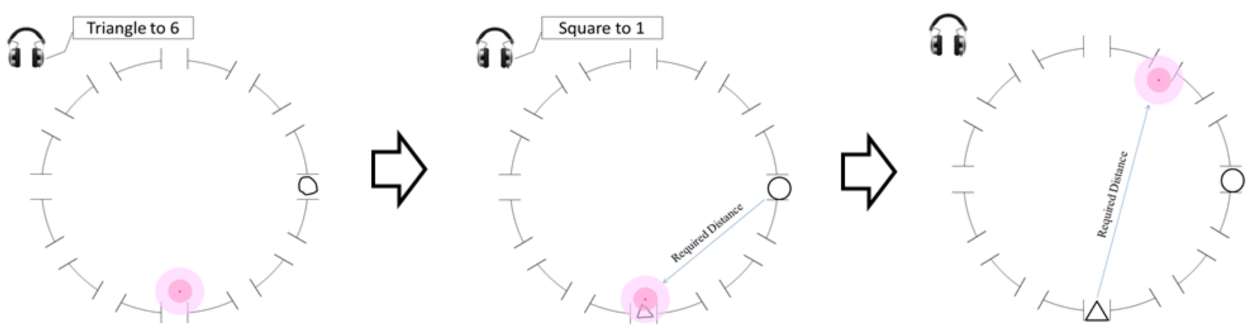


Figure 3-5: An example segment of the progress of clock filling task for the first part of the experiment. Note that the instructions contain both the location and the type of sketches.

the next number were issued through headphones. The places of numbers on a wall clock are simple and well known to everybody. Any misrecognition that occurred on the wall clock also became contextually inferable since they contradicted with the rest of the clock.

Hence in this part we collected gaze behaviors during the interaction with contextually inferable sketch misrecognitions. Figure 3-4 shows how clock filling procedure progressed.

3.3.2 Creating Contextually Non-conflicting Misrecognitions

In the second part, the participants were required to place basic geometric shapes (circle, square and triangle) on the clock template instead of numbers. Again the announcements specified the place and type of each sketch. The announcements were triggered by the participants' first stroke also at this part. Figure 3-5 shows how the task of filling the clock templates with shapes progressed.

The participants built scenes consisting of random geometric shapes on which they had no prior familiarity. Hence there was no contextual coherence in any of the scenes for the participants. Any misrecognition was non-conflicting with the context being concealed within other objects.

3.4 Procedure

We collected data from 12 people (9 males, 3 females) between the ages of 23 and 29. Participation was voluntary and the participants were free to leave anytime. Each session lasted approximately 30 minutes. We did not inform the participants that sketch recognitions were predetermined in order to preserve natural human–computer interaction.

Each participant went through a calibration process for the eye tracker. We did not use a chin rest but asked participants to preserve the position they had during the calibration throughout the experiment as much as possible.

After calibration the participants completed two tasks in order to get familiar with the sketching surface and the setup. The first task involved drawing the path between the entrance and the exit of a labyrinth given with top view. The second was to complete a connect-the-dots puzzle. Then they went through a trial stage of completing a wall clock with numbers. At this stage we made sure that they fully understood the task. Then the participants repeated the task 10 times for both parts of the experiment. Each participant filled 20 clock templates in total.

We administered a post–experiment questionnaire to see if the users believed that they were working on a real sketch recognition interface. We asked participants whether they noticed that there were any outside effects on the responses of the interface. All participants said that they did not suspect that the sketch recognition module was real and they were interacting with a real sketch recognition interface.

Chapter 4

FACTORS ON CORRECTION PERFORMANCE

In this chapter we present quantitative analysis on participants' sketch misrecognition detection performance and how this performance is affected by contextual inferrability and visual field of sketch misrecognition.

We asked participants to mark sketch misrecognitions by drawing a line over them when they notice them during the experiment. There were sketch misrecognitions which participants left unmarked, implying that these were not noticed. We use the ratio of unmarked sketch misrecognition count to total sketch misrecognition count as the measure of the noticing performance.

4.1 Effect of Sketch Misrecognition's Contextual Inferrability

Figure 4-1 shows the distributions of realized/unrealized sketch misrecognitions when sketch misrecognitions were contextually *inferable* or *non-conflicting*. For the ease of reading we will refer contextually inferable sketch misrecognitions as CI and non-conflicting misrecognition as CN. Each bar represents all CI (acquired in the first part of the experiment) or CN (acquired in the second part of the experiment) misrecognitions occurred. The dotted parts represent the ratio of noticed misrecognitions while the lined parts of the bars show the ratio of missed sketch misrecognitions. The distribution on the left shows the noticing rate when the sketch misrecognitions are inferable as the distribution on the right shows the noticing rate when they are non-conflicting. Distributions show that users are more likely to notice sketch misrecognitions when the misrecognition is CI and this difference was found to be significantly different (p<0.01) with ANOVA test.

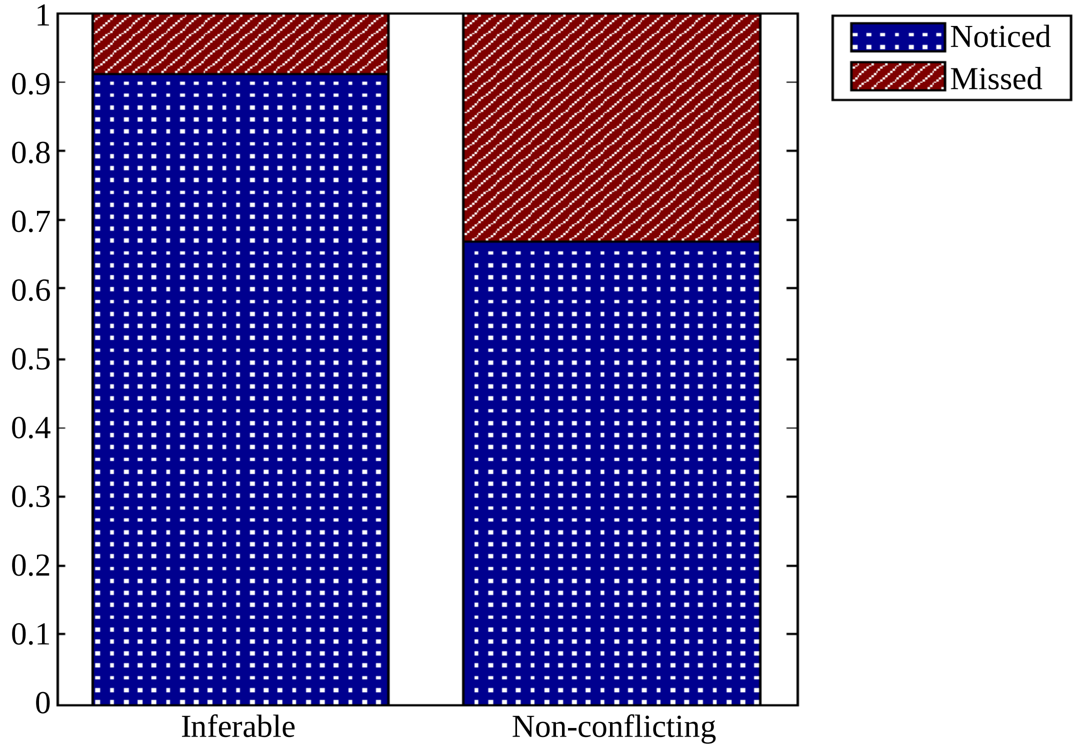


Figure 4-1: The distribution of noticed and missed sketch misrecognitions by the relation of the misrecognition with the context (inferable of non-conflicting).

This result coincides with the previous work on change detection. When the changes made to the objects' place or type in a static scene is contradictory with context, people are able to notice these changes more easily. Hence this analysis shows that the act of sketching does not enrich the humans' visual representation of the scene enough to notice any non-conflicting change as good as an inferable one.

4.2 Effect of the Sketch Misrecognition's Visual Field

The experiment interface gave sketch misrecognition stimuli at different distances from the gaze. Using these measurements we grouped sketch misrecognition stimuli by their place in the human visual field and investigated how participants' performance for noticing misrecognitions changes in respect to the misrecognitions place in visual field.

Human visual field is divided into 3 main parts which have different visual qualities. At the center of human visual field is foveal area in which we have clear vision and see objects in detail. The most outer part of our visual field is peripheral area. The vision gets more blurry as visual field goes from foveal area to peripheral area. There is also a transition field called parafoveal area between foveal and peripheral area. Figure 4-2 shows these fields.

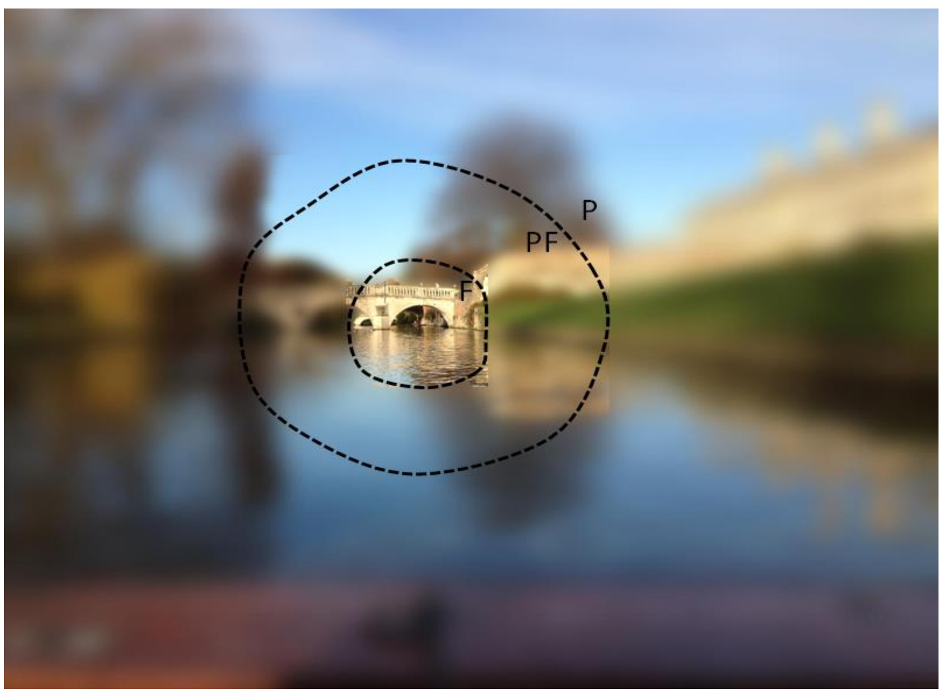


Figure 4-2: Visual areas of the human visual field. F denotes the Foveal (1^o-2^o), PF denotes the Parafoveal (2^o-6^o) and P denotes the peripheral (6^o-220^o) parts of human visual field. Note that the visual area borders are slightly irregular.

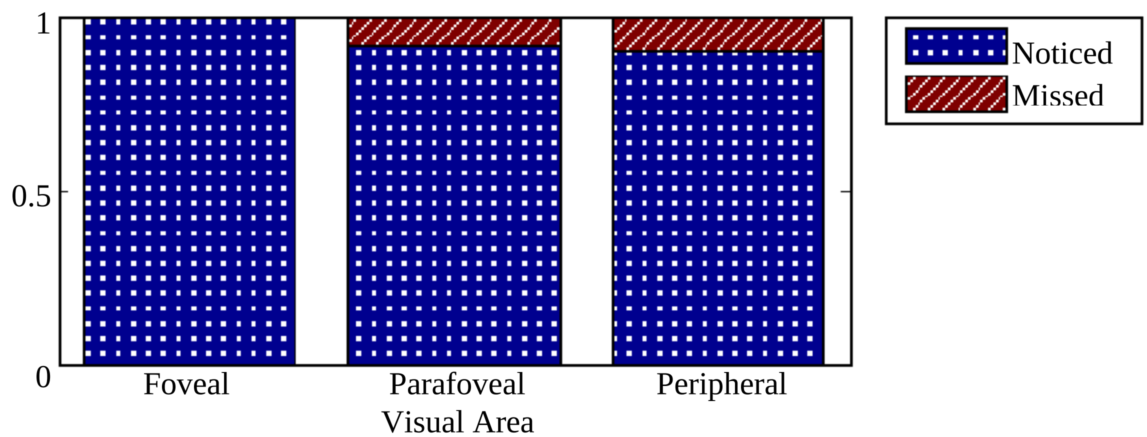


Figure 4-3: Correction rates of sketch misrecognitions in different visual fields when they are contextually inferable

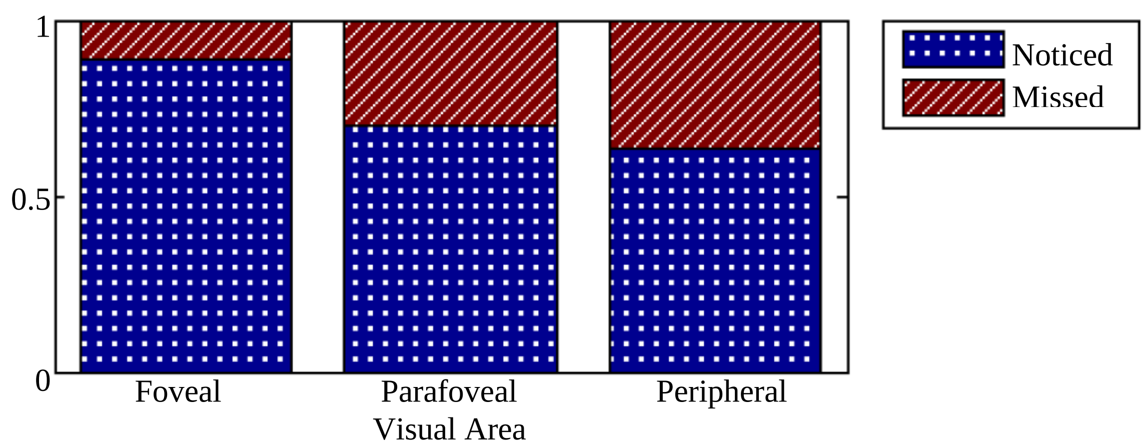


Figure 4-4: Correction rates of sketch misrecognitions in different visual fields when they are contextually non-conflicting

We hypothesized that as the angular distance of the visual field becomes smaller the rate of noticed misrecognitions would go higher. Figure 4-3 and Figure 4-4 show the noticing rates of sketch misrecognitions in different visual fields when the sketch misrecognitions are CI and CN respectively. Again the dotted parts show ratio of noticed and the lined parts show ratio of unrealized sketch misrecognitions. Noticing rates show a decreasing trend as the stimulus moves towards periphery for both cases. When the sketch misrecognition is CI, visual field and realization performance have a strong correlation of -0.94. This correlation slightly increases to -0.96 when the misrecognition is CN. These findings are also consistent with previous research on the effects of the distance on scene change detection performance [15], [16].

Chapter 5

FEATURES

In this chapter we first present our observations on participants' gaze behavior as they moved on to correct a sketch misrecognition. These observations provide an insight to correction intention gaze behavior. Then we present the feature set and explain how we constructed this set. We give the intuition for each feature.

5.1 Gaze Behaviors

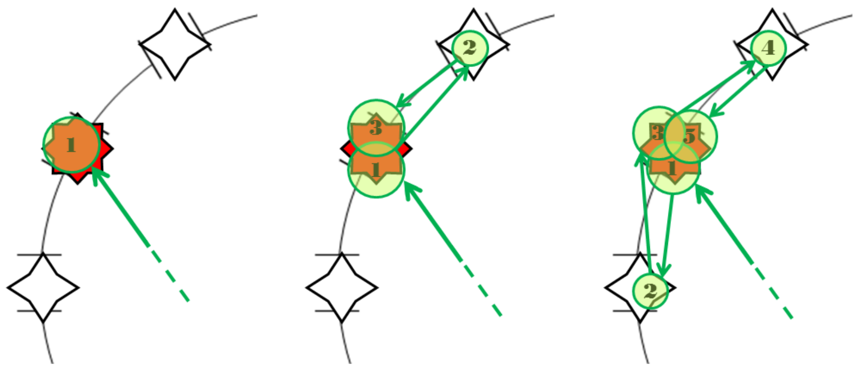
We observed that there are three distinctive gaze behaviors common among all users prior to the initiation of the act of correcting sketch misrecognitions. Figure 5-1 shows gaze trajectories for all behaviors.

Direct Comeback: Users' gaze comes back directly to the misrecognition and stays on until they mark it.

Double Take: Users do a double take after the first fixation to the misrecognition before they mark it as a misrecognition.

Investigation: Users make fixations to other related sketch recognitions until they mark the misrecognition.

Direct comeback behavior occurs regardless of the misrecognition being contextually inferable or non-conflicting. On the other hand double take and comparison behaviors occur particularly when sketch misrecognitions are CI. Direct comeback gaze behavior is also the most frequent behavior, both when misrecognitions are CI and CN. Figure 5-1 shows how gaze moves for each of these behaviors.



a b c

Figure 5-1: Example gaze paths for (a) Direct Comeback (b) Double take and (c) Investigation behaviors. The green circles show fixations. The radius of the circles is proportional to the duration of the fixation. The red 8-point stars stand for sketch misrecognition and white 4-point stars denote correct sketch recognitions close to the sketch misrecognition.

When the sketch misrecognition is CI the users have more means to decide whether a sketch recognition feedback is incorrect. To be able to confirm the misrecognition they examine other sketch recognitions on the scene and collect clues indicating that there should have been another object in the place of the sketch misrecognition. Then they confirm the misrecognition and perform correction. Since contextually non-conflicting misrecognitions are coherent with the context, investigating the scene does not provide users information. Hence investigating behavior is not observed for contextually non-conflicting sketch misrecognitions. When the sketch misrecognition is non-conflicting with the context, the users depend on the memory of their sketch to conclude the suspicion. This creates mostly direct comeback behavior. It might be the case that once the user gaze comes back to the sketch misrecognition, it can only wait there while the information attributed to the sketch is retrieved from the user's memory.

Double take is generally regarded as an act of surprise. We believe that this behavior is the result of the surprise effect that a CI sketch misrecognition have on users.

This behavior pattern suggests that for realizing CN sketch misrecognitions, people depend on their visual memory. On the other hand for CI, they can take advantage of context information which manifests itself as double take and investigation behaviors. Taking the results in Chapter 4.1 into account we can say that, sketching does not contribute to visual memory enough to make up for the advantages of contextual clues.

5.2 Preprocessing

Our data consisted of complex gaze patterns. Hence we segmented gaze data into fixations and saccade for simplification. Fixations are characterized by the lower speed of gaze compared to saccades. Hence we used acceleration and velocity of eye gaze as features [43] and trained a K-Means Hidden Markov Model [44] for differentiating fixations from saccades. We used JaHMM[1] library in java language for training K-Means HMM. We merged successive fixations that are closer to each other than 1^{o}, since during these fixations there are no changes to users' foveal area (1^{o} – 2^{o}). We also constructed *visits*. A visit to an object starts when the gaze enters and ends when the gaze exits the objects area. Visits can contain multiple successive fixations on an object. We attached fixations and visits located inside the borders of an object to that object.

5.3 Features

We find it necessary to explain how we approached the classification problem in order to convey how we designed our features and what they mean. Users made fixations to all sketch misrecognition that they corrected. However they also made fixations to correct sketch recognitions with non-correction purposes. Our problem was to distinguish the

[1] JaHMM library can be accessed from http://www.run.montefiore.ulg.ac.be/~francois/software/jahmm/

correction intending fixations from general purpose fixations to sketch recognitions. In order to understand the intention behind a fixation we took into consideration not only the fixation itself but previous gaze movements regarding the sketch recognition. Hence we analyzed all gaze movements starting from the occurrence of sketch recognition (stimulus) until the fixation to this sketch recognition. We calculated features over this data window.

We picked features depending mainly on research on psychology and our observations on the data. A list of feature names and explanations are in Table 5-1.

Table 5-1: Features

Abbreviation	Feature Name	KS Statistic	P(<)
DurLastFix	Duration of the Last Fixation	0.6510	1.2e-131
NofFix	Number of Fixations	0.1370	1e-6
MeanDurFix	Mean Duration of Fixations	0.5587	4.2e-97
DurLastVisit	Duration of the Last Visit	0.7163	2.7e-159
NofVisits	Number of Visits	0.1953	1e-6
MeanDurVisit	Mean Duration of Visits	0.5970	9e-111
FixVisit	Fixations per Visit	0.1116	2.8e-4
MeanSpan	Mean Span of Approaching Saccades	0.1023	1.2e-3
FirstSpan	Span of the First Approaching Saccade	0.1532	1e-7
LastSpan	Span of the Last Approaching Saccade	0.0977	2.2e-3
MeanSpeed	Mean Speed of Approaching Saccades	0.0959	2.9e-3
StdFixTime	Standard Deviation of Fixation Start Times	0.3536	3.85e-39
StdVisitTime	Standard Deviation of Visit Start Times	0.3517	1e-39
TimeStimFF	Time Between Stimuli and the	0.2052	1.8e-13

	First Fixation		
TimeFFCor	Time Between the First Fixation and Correction	0.3403	2.7e-36
NofSketchElse	Number of Sketches made Elsewhere During a Fixation	0.0443	2.7e-3
NofStrokesBtw	Number of Strokes Between Two Fixations	0.2949	1e-6
NofVisitOther	Number of Visits to Other Sketch Recognitions Between Two Fixations	0.2863	1e-6
VisitDurOther	Duration of Visits to Other Sketch Recognitions Between 2 Fixations	0.2865	7.6e-26
NofFixEmpty	Number of Fixations to Empty Space Between 2 Fixations	0.3192	1e-6
FixDurEmpty	Duration of Fixations to Empty Space Between 2 Fixations	0.3353	3.1e-35

Research on change detection [45] proved that people fixated longer to changes on a scene. Hence we used features based on fixation duration such as DurLastFix, MeanDurFix, DurLastVisit and MeanDurVisit. Also both fixation duration [30][31][34] [36][38] and frequency [34][46] were used for inferring users interest in previous studies. We also hypothesized that sketch misrecognition ignite more interest than correct sketch recognitions. Hence we extracted values indicating fixation frequency and count such as NofFix, NofVisits, FixVisit, StdFixTime and StdVisitTime. We also used distance related features such as MeanSpan, FirstSpan, LastSpan and MeanSpeed with the intuition that gaze would leave its trajectory when the person realizes something that attracts attention. [46] TimeStimFF and TimeFFCor are features about timing of fixations and corrections. We hypothesized that people would realize and react to sketch misrecognitions faster. Finally we included features denoting distraction from the sketch recognition such as NofSketchElse, NofStrokesBtw, NofVisitOther, VisitDurOther, NofFixEmpty and FixDurEmpty based on our observations.

5.4 Distribution of Features

We have investigated how different the distributions of features for correction intending and general purpose fixations are. We expected that the fixation duration and frequency based features would be more likely to take higher values prior to correction intending fixations than general purpose fixations. We also expected that distance based features including saccade speed would take higher values for correction intending fixations since gaze would move faster and from a greater distance when noticed something unexpected. Timing based features; TimeStimFF and TimeFFCor would be more likely to take lower values since reaction time to a sketch misrecognition would be shorter and correction would take place sooner. Intuitively, features indicating distraction from the sketch recognition would also take lower values for correction intending fixations since once the users fixated on a sketch misrecognition they would first correct it before attending to something else.

We calculated empirical cumulative probability distributions for each feature for correction intending and general purpose fixation classes. We compared distributions from both classes. The results support our intuitions about how feature values would change based on class. The plots for probability distributions for each class are in APPENDIX.

To see if the differences between distributions are significant, we applied Kolmogorov - Smirnov test. Kolmogorov – Smirnov test is widely used to determine if two empirical cumulative probability distributions are significantly different. It takes the maximum value of difference between two distributions and checks this value for significance. The test is originally designed for continuous values. However there are extensions for distributions of discrete values. We used the extension by Altavela [47]. All features proved to have significantly different distributions for two different classes. KS statistic and p values are presented in Table 5-1. Note that even though small enough for significance, the p values for MeanSpan, FirstSpan, LastSpan and MeanSpeed are noticeably higher than other p values.

Chapter 6

CLASSIFICATION

In this chapter we give detailed description of the machine learning framework we built for discriminating correction intending gaze behavior. We explain how contextual inferrability and visual field of sketch misrecognitions affect classification accuracy. Then we demonstrate how accuracy of intention detection changes as we go back in time starting from correction. Finally, we present analyses on how different feature groups contribute to classification accuracy.

6.1 Classification Framework

As we explained in previous chapter our problem was of distinguishing fixations with correction intention and general purpose. We labelled each fixation to a sketch misrecognition at the correction time as correction intending. If the fixation extended beyond the pen down, which indicates also start of correction act, we divided the fixation from the pen down and used only the part prior to pen-down, because our goal is to detect correction intention before users take action themselves. We took gaze data from stimulus to pen down and extracted features. On the other hand, each fixation to a correct sketch recognition was labeled as general purpose fixation. We took the data from stimulus to the end of each general purpose fixation and extracted features for this class. We acquired 381 correction intending fixation instances and 4582 general purpose fixation instances. The reason for this imbalance to begin with is that sketch misrecognitions were sparser than correct sketch recognitions during the experiment. Also we were able to acquire one correction intending fixation for each sketch misrecognition, while there were mostly more than one general purpose fixations for correct sketch recognitions.

After we created feature vectors we built the machine learning framework. We used with Support Vector Machines (SVM) with radial basis kernel function. We used LIBSVM library [48]. We performed 10 shuffles with 5 fold cross–validation. Because the sample counts in different classes are unbalanced, we picked random samples from the majority class as many as the number of samples in minority class, in each shuffle. Hence we had a baseline accuracy of 50% to compare our accuracy results more appropriately. We used 80% of the data for training and the rest for testing.

We acquired 86% classification accuracy with 3.15% standard deviation. True positive and negative rates and Type I and II error rates are in Table 6-1. True positive and negative rates are close to each other which imply that the classifier accuracy does not come from prediction of only one of the classes. The system detected correction intending and general purpose gaze in balance.

Table 6-1: True positive, true negative, false positive and false negative rates

		Test Outcome	
		Correction intending fixation	General purpose fixation
True Label	Correction intending fixation	86.56%	13.44%
	General purpose fixation	14.43%	85.57%

False positive (general purpose fixations labeled as correction intending) rate is an important factor that may affect user satisfaction while using this system. Because this rate shows how many unnecessary correction-assistances the system will give. Unnecessary correction-assistance may create a disturbance for the user. Hence we need to conduct an experiment to evaluate this system and see how much a false positive rate of 14.43% will disturb users' interaction. There is also the possibility of users thinking that the sketch recognition is false even if it is true.

6.2 Effect of Background Information on Classification

We investigated how much contextual inferrability or non-conflicting nature of sketch misrecognitions effect classification. For this purpose, we trained SVM only including the data from the CI and CN sketch misrecognitions respectively. We tested each model again with CI and CN data. We present the results in Table 6-2. Note that ANOVA test on the values in Table 6-2 showed that all values are significantly different from each other.

Table 6-2: Accuracy results for CI and CN data sets

		Test	
		CI	CN
Train	CI	83.2%	86.3%
	CN	78.9%	88.7%

When we train with CI data and test with CN data we get a higher accuracy than the testing with CI data. We believe that the reason for this is that the range of the behaviors observed during correction intending fixations in CI data is more diverse than CN data. Thus when we test a versatile model with a more limited data, we acquire a higher accuracy than testing with the data of matching versatility. A similar notion goes for the case when we train the model with CN data. When we test this model with CI data we get lower accuracy than testing with CN data. Hence when we test a limited model with versatile data we get lower accuracy than testing with data limited in the same way. An interesting point here is that when we train and test with CN data we get higher accuracy than training and testing with CI data even though in these cases both training and testing data have same diversity. The reason for this may be that CN data is simpler and some of our features are more successful at representing the difference between correction intending and general purpose fixation classes in this case. On the other hand in CI data, gaze patterns may get more complicated and our features are not as successful as for the CN data.

6.3 Effect of Visual Area of Stimuli on Classification

We investigated how much the visual field in which the sketch misrecognition is located affects classification accuracy. Hence we used each correction intending class sample as a test sample a number of times and investigated the correlation between the correct classification rates of samples with their visual field. For this purpose we divided correction intending data randomly into 5 groups. We used in turn each group as test set and the remaining groups as training set. We repeated this random division 10 times so that each sample would be in different training and test groups. For each test set in each division we also picked a new set of random general purpose fixation samples in the number of correction intending fixation samples 10 times and performed training and testing. Hence we used every correction intending fixation sample as a test sample 100 times. We calculated the rate of correct labels among 100 labels and acquired correct classification rates. Figure 6-1 shows mean correct classification rates with standard deviation for sample groups by their visual field.

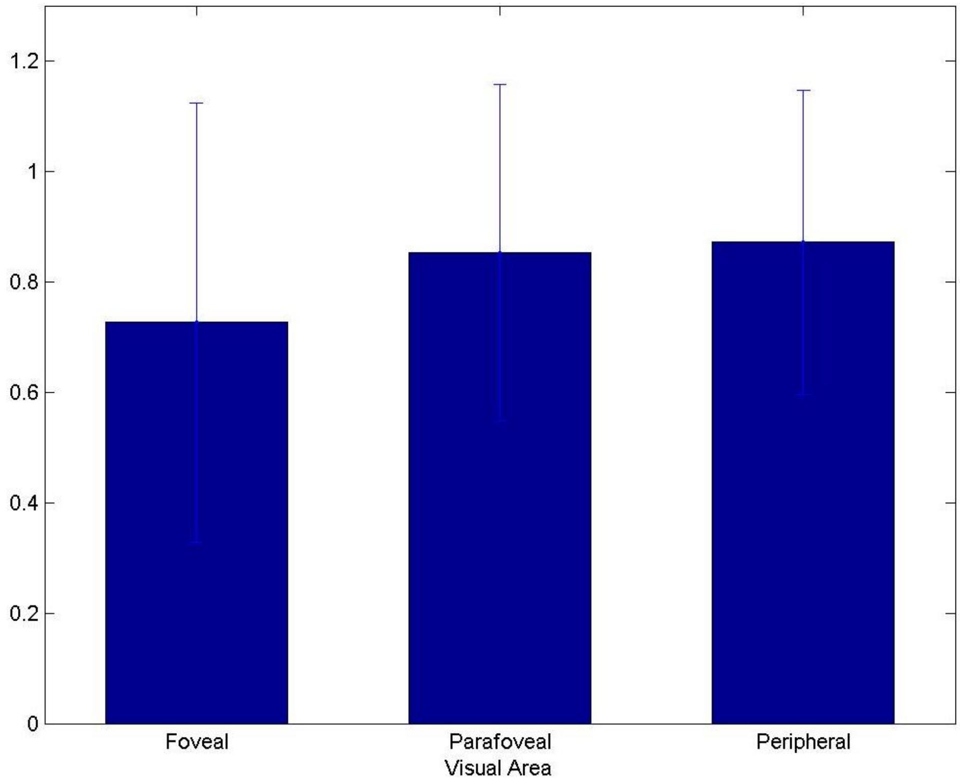


Figure 6-1: Change of correct classification rate by the visual area. The lines on top of each bar denote the standard deviation of correct classification rates of samples in each visual area.

The correlation between sketch misrecognitions' visual field and their correct classification rate is 0.95. The reason may be that as the gaze moves away from the misrecognition, it gets harder to notice it. When the user finally realizes the misrecognitions there is a surprise effect that causes the gaze patterns to be more characteristic.

6.4 Prediction as a Function of Time

The sooner the system detects sketch misrecognitions the more efficient user – computer interaction becomes. Hence we investigated how earlier from users' reaction can our system detect sketch misrecognitions. Figure 6-2 shows the classification accuracies at the times preceding the user's correction. For example time "t" denotes the end time of data frames acquired as explained in section 5.3. Vertical axis gives the classification accuracy acquired at this moment. Time "t-500" denotes the moment 500 milliseconds before the correction or fixation end time. Again the vertical axis value for this point gives the classification accuracy using the features extracted from the data sample between stimulus and this time.

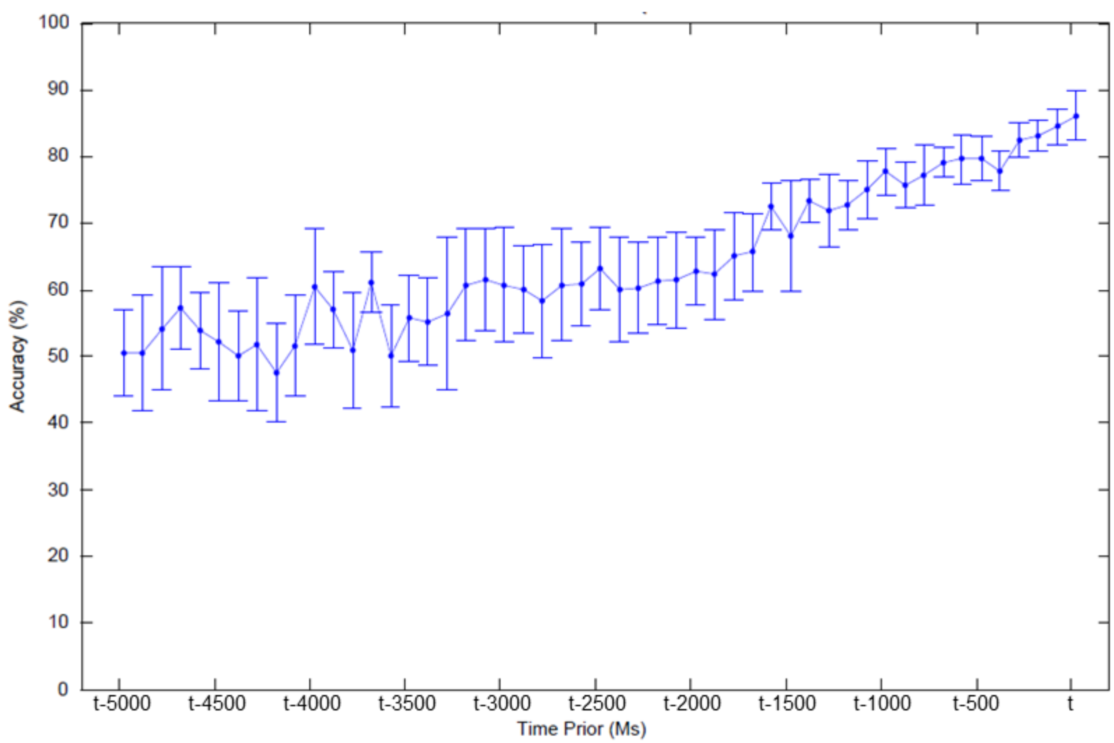


Figure 6-2: Change of prediction accuracy by prediction time prior to correction

The classification accuracy is around %75 until 1 second before the correction. Accuracies decrease slightly steeper after 1 second. We see that there is a tradeoff between classification accuracy and the time advantage provided to the user. If we want the users to gain time advantage then we have to take the risk of incorrect predictions. If we want to make the prediction as accurate as possible then we should wait until the last moment to make the prediction.

6.5 Feature Contribution

Another point to investigate was how each feature and their interactions with others contributed to classification accuracy. One way to observe this was calculating accuracies for every possible feature combination. However this approach was computationally expensive and even if we abided this expense, we would not be able to understand which features and in which combinations did a significant contribution. Therefore instead of using each feature individually, we grouped features semantically and investigated each groups' contribution. Table 6-3 shows semantic grouping of features given in Table 5-1.

Table 6-3: Feature groups

Duration Group	Frequency Group	Distance Group
DurLastFix	NofFix	MeanSpan
MeanDurFix	NofVisits	FirstSpan
DurLastVisit	StdFixTime	LastSpan
MeanDurVisit	StdVisitTime	Mean Speed
TimeStimFF	NofStrokesBtw	
TimeFFCor	FixVisit	
VisitDurOther	NofSketchElse	
FixDurEmpty	NofVisitOther	
	NofFixEmpty	

We calculated accuracies for all 7 combinations of these groups with 5 fold cross - validation and 10 shuffles. We obtained accuracies to represent the 8th combination (no features case) by flipping a virtual fair coin. Table 6-4 shows each combination and how accuracy result for each combination was acquired.

Table 6-4: Feature group combinations. ‘0’ indicates that the feature group is absent; while ‘1’ indicates that the feature group exists in the combination

Combination			**Accuracy acquired by**
Duration	**Frequency**	**Distance**	
0	0	0	Virtual Coin Flip
1	0	0	10 shuffles with SVM
0	1	0	10 shuffles with SVM
0	0	1	10 shuffles with SVM
1	1	0	10 shuffles with SVM
1	0	1	10 shuffles with SVM
0	1	1	10 shuffles with SVM
1	1	1	10 shuffles with SVM

We tested each feature group’s effect on classification accuracy using 3-way ANOVA. Each feature group was used as a factor with 2 levels; existence (taking the value of 1) or absence (taking the value of 0) in the model.

Figure 6-3 shows the marginal mean change created by single factors. The biggest change is created by Duration feature group which increases the classification accuracy significantly from 60.3% to 84.3%. Frequency feature group also creates a significant increase in the classification accuracy though not as high as Duration group. On the other hand Distance group also increases classification accuracy but this increase is not significant.

We investigated how pairs of features groups interacted with each-other. Hence we inspected how the effect of a feature group on accuracy changes when the level (existence) of another group changes.

Figure 6-4 shows the interaction between Duration and Frequency group. It shows that when the Duration group is included in the model the classification accuracy increases

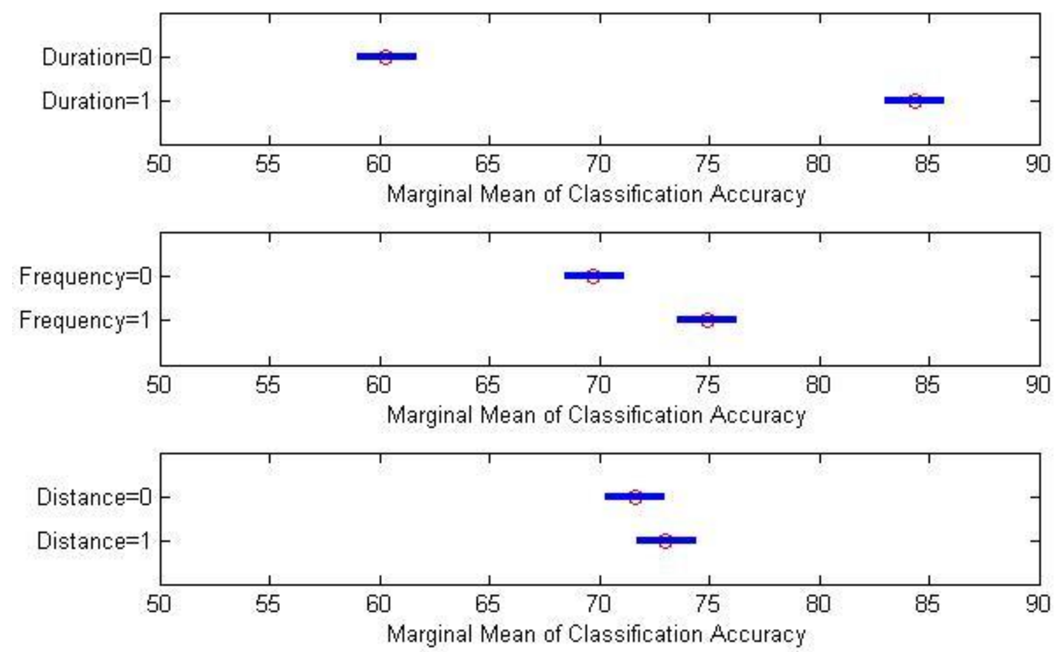


Figure 6-3: Marginal mean changes for each feature group. Red circles denote the marginal means while blue lines denote 0.95 confidence intervals. Overlapping of the confidence intervals of two values indicates that the values are not significantly different from each-other. The values are significantly different if their confidence intervals do not overlap.

significantly irrespective of whether the Frequency group is included or not. When we add Frequency group to the model when the Duration group is absent it makes a significant contribution to the classification accuracy. On the other when the Duration group is already in the model adding the Frequency group does not cause a significant change in the accuracy. Hence we can say that even if the Frequency group is informative on the differences of two classes the information it contain does not surpass Duration group. There are almost no samples that were classified incorrectly with the Duration group only and were able to be transferred to the correct class with the information coming specifically from the Frequency group.

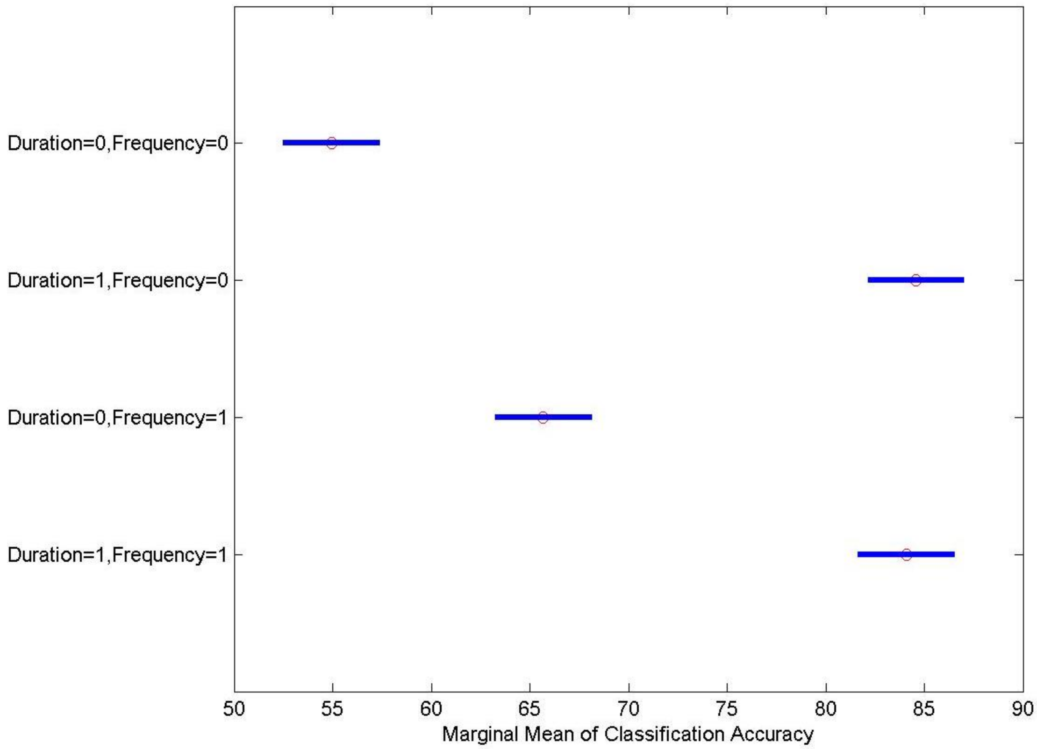


Figure 6-4: Effect of interaction between Duration and Frequency feature groups on classification accuracy

Figure 6-5 shows the interaction between Duration and Distance feature groups. The same rule that exists for Duration-Frequency group combination goes also for this combination. Similar to the Frequency group the Distance group does not contribute any discriminative information on the two classes to the Duration group. However, differently from the Frequency group, Distance group's existence in the Duration group's absence does not increase the classification accuracy. We can say that Distance group contains hardly any information on the difference of the two classes.

Figure 6-6 shows the interaction between the Frequency and the Distance groups. In the absence of Distance group adding the Frequency group creates a significant increase in the accuracy while in the existence of Distance this increase is insignificant due to Distance group's contribution to the accuracy when the Frequency group is unused. However Distance group has no effect on the final accuracy where both groups were used.

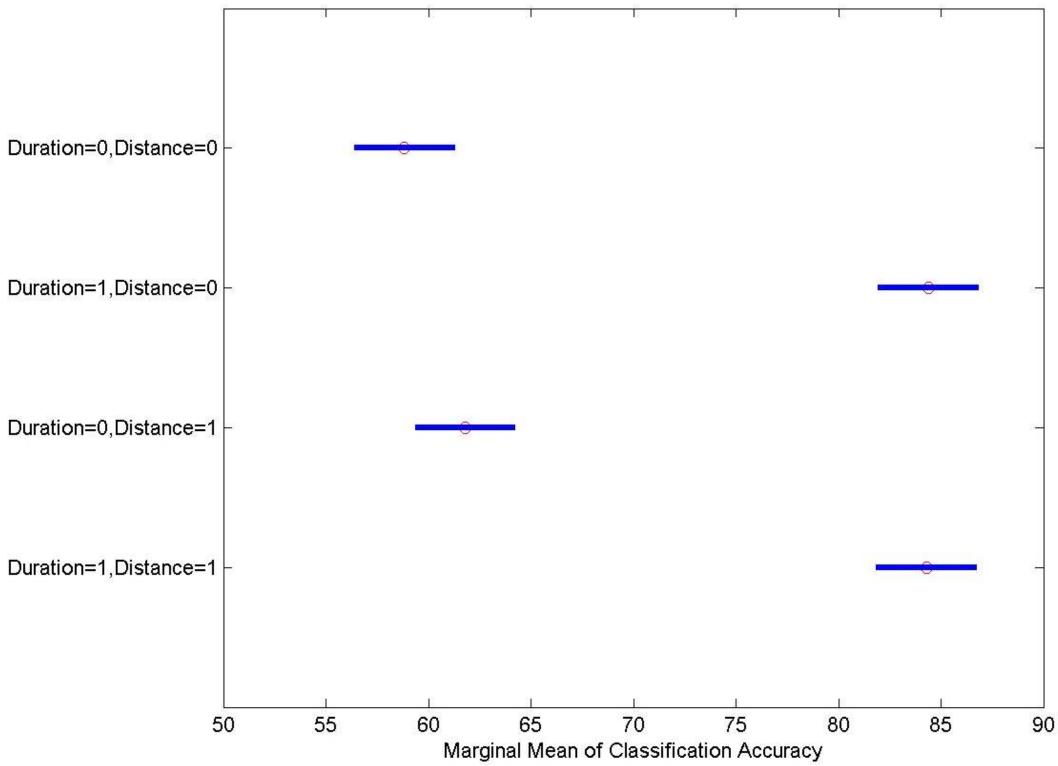


Figure 6-5: Effect of interaction between Duration and Distance feature groups on classification accuracy

The analyses show us that the Duration based features are the most informative among others, single handedly achieving classification accuracy around 85%. Frequency group could have been useful in the absence of Duration features however they are redundant when combined with Duration features. Distance feature group has no contribution to the accuracy either on its own or combined with any other feature group. Even though Kolmogorov-Smirnov test concluded that distributions of Distance features for two classes were significantly different, remember that the test also gave the highest p-values for significant difference.

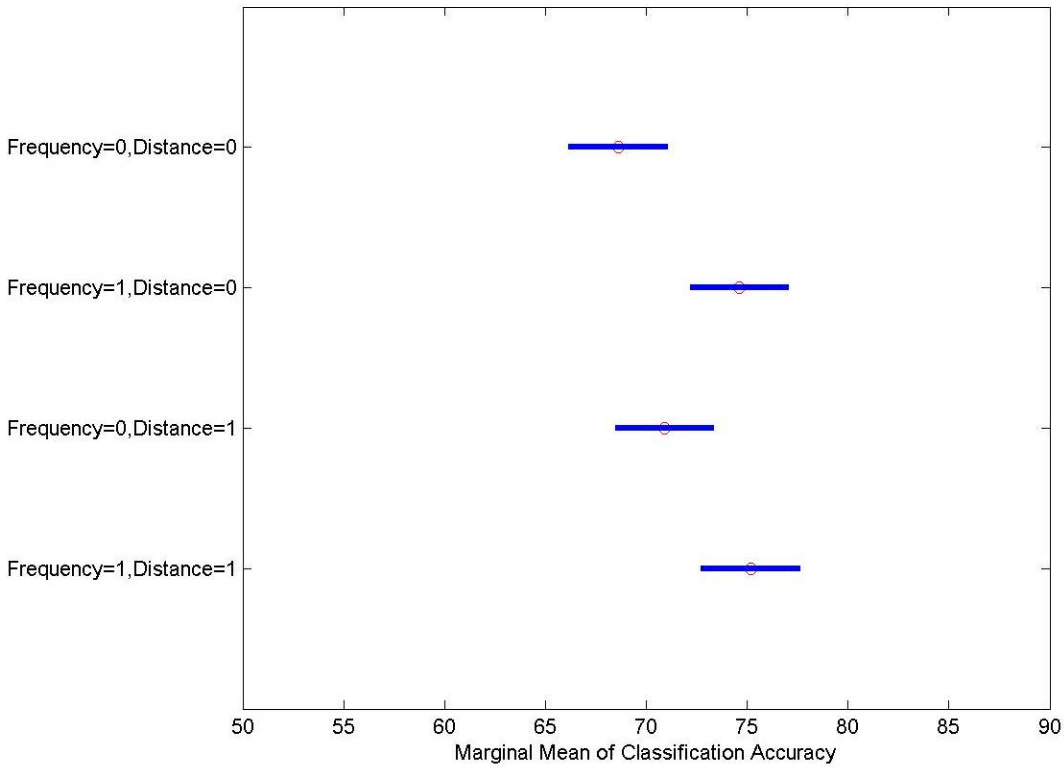


Figure 6-6: Effect of interaction between Frequency and Distance feature groups on classification accuracy

Chapter 7

CONCLUSION AND FUTURE WORK

In this study we proposed a framework for detecting users' decision on a sketch recognition's correctness by examining their eye gaze. We conducted a Wizard of Oz experiment for collecting eye gaze patterns under two sketch –based interaction conditions; sketch misrecognition's visual field or them being contextually inferable or non-conflicting. We showed that users' eye gaze movements display distinctive characteristics reflecting the natural reaction to misrecognitions. We extracted a set of features that represent these characteristics of users' eye gaze behavior. Finally, using the feature set we build a machine learning framework to recognize user's intention to correct sketch misrecognitions. The system achieves 86% prediction accuracy. This high accuracy value indicates that this framework can be realized as a system that eases the burden of fixing the sketch misrecognition on users. It also shows that eye gaze, especially with its proactive properties, can be used as a supporting modality for increasing the efficiency of human-computer interaction, specifically pen based interaction.

We showed that when the sketch misrecognition is incoherent within the scene context then there is a higher chance that the users detect this misrecognition. Users also tend to miss sketch misrecognitions as the misrecognition moves from inner visual fields to outer visual fields. These results coincides with the results acquired from studies on gaze behaviors on static (no sketching) scenes, indicating that sketch action has no substantial contribution to human's visual representation.

We also investigated how much time advantage the system can provide to users before they take action themselves. The results showed a trade-off between the reliability of the prediction and the time advantage. The system provides 77.7% accuracy 1 second before the users initiate action. The accuracy was higher than chance as much as 2 seconds prior to the user action.

We analyzed which properties of the eye gaze provided information to differentiate gaze behavior as the users detected misrecognitions. Hence we grouped the features into semantically meaningful subgroups; duration, frequency and distance. Among these 3 groups the features indicating the duration of the gaze proved to be the significantly more informative than other two. Distance group had no significant contribution to prediction accuracy while frequency group contained enough information to raise the accuracy from baseline significantly on its own and make a little contribution to duration group though not significantly. These results indicate that total duration of the gaze spent on sketch misrecognition has substantial information regardless of how segmented or how distant the fixations are.

We will investigate how different sketch recognition feedback mechanisms affect gaze behavior and users' performance to catch sketch misrecognitions. We will search for feedback mechanisms that will allow users to realize contextually non-conflicting misrecognitions with a higher performance.

We also plan on extending the current framework to be able work in real time and building a real time support system for misrecognition correction. We will investigate whether the users adapt and change their gaze behavior knowing that the interface will give support reading their intention. Taking this possible factor into account we plan on develop different support strategies and evaluate these with user studies.

APPENDIX

This appendix presents probability and cumulative probability distributions of values belonging to two different classes for each distribution. Distributions are acquired empirically. Also Kolmogorov-Smirnov (KS) Statistics is given for each feature. These probability distributions provide further understanding on how feature values have different trends for different classes. The green lines with circle markers represent probabilities (either cumulative or not) for feature values extracted from correction intending gaze behavior, while red lines with triangle markers show probabilities for the same feature extracted from other gaze behavior sketch misrecognitions.

DurLastFix: Duration of the Last Fixation (Kolmogorov-Smirnov Statistic= 0.6510)

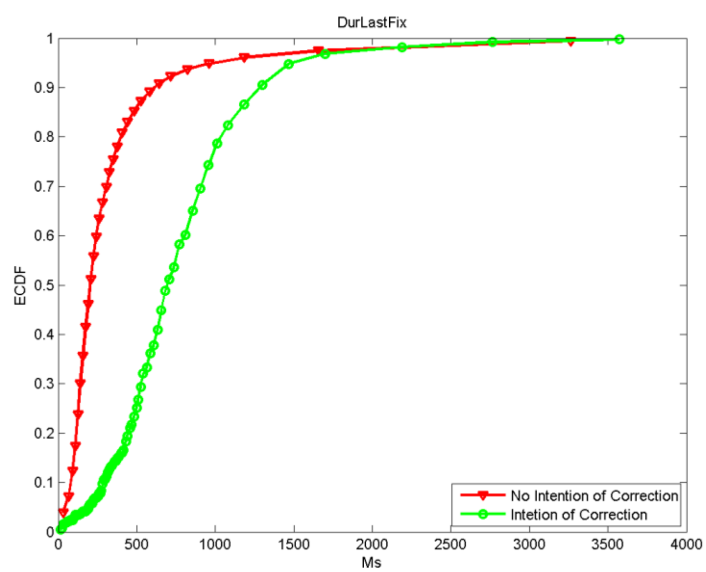


Figure A. 1: Empirical cumulative probability distributions for DurLastFix

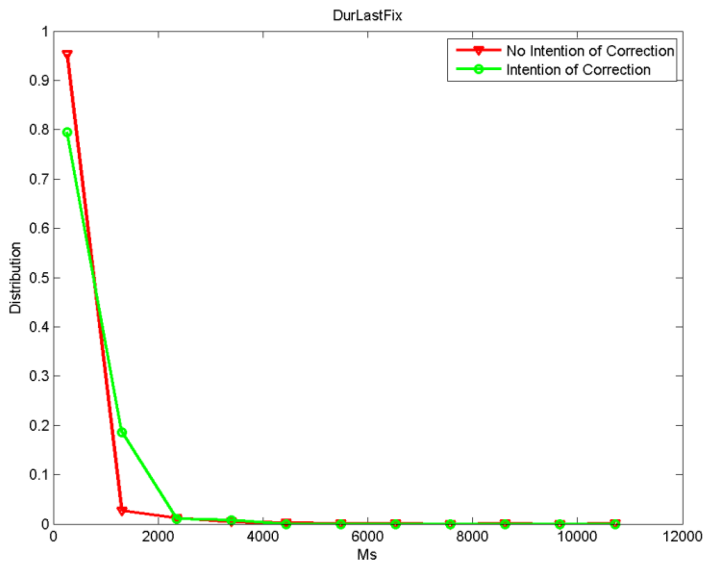


Figure A. 2: Empirical probability distributions for DurLastFix

NofFix: Number of Fixations (Kolmogorov Smirnov Statistic=0.1370)

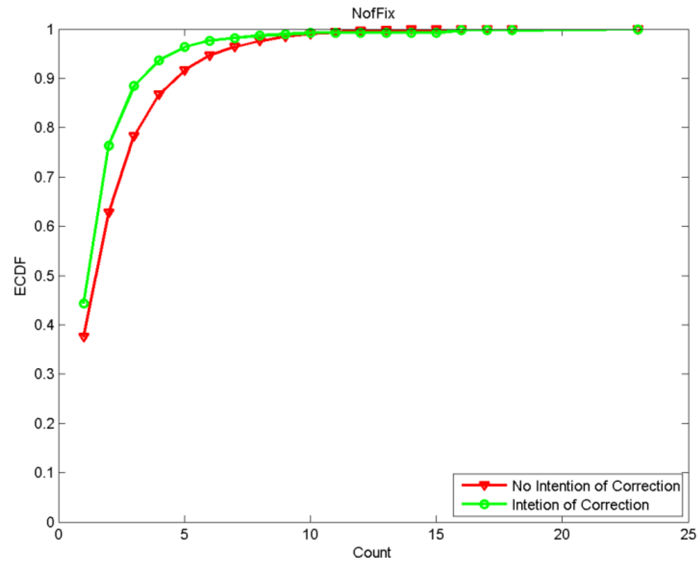


Figure A. 3: Empirical cumulative probability distributions for NofFix

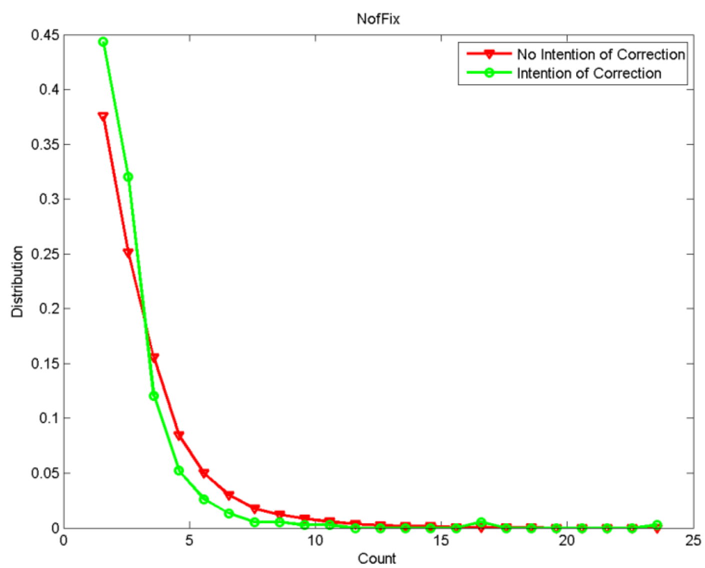


Figure A. 4: Empirical probability distributions for NofFix

MeanDurFix: Mean Duration of Fixations (KS Statistic=0.5587)

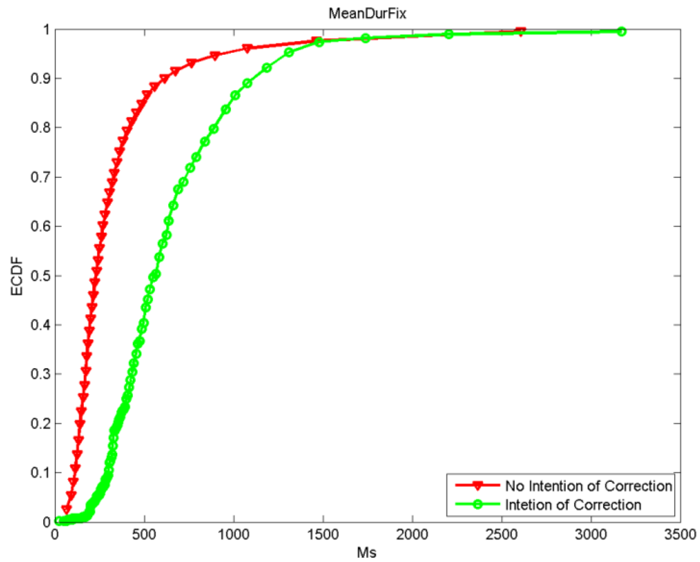


Figure A. 5: Empirical cumulative probability distributions for MeanDurFix

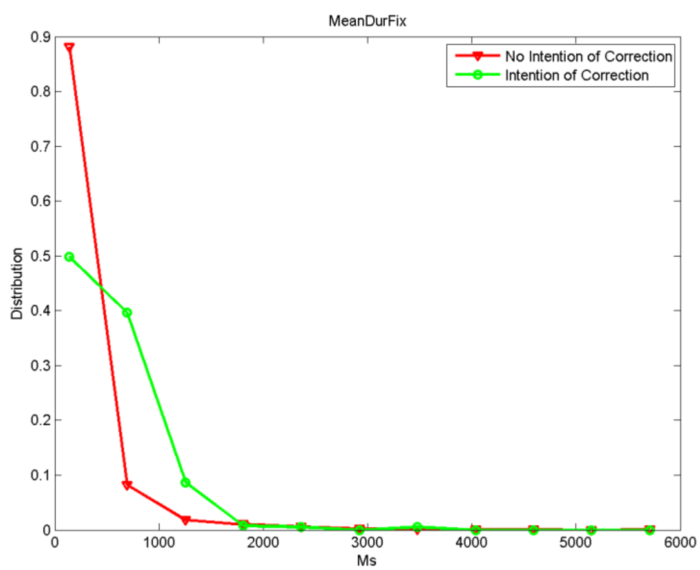


Figure A. 6: Empirical probability distributions for MeanDurFix

DurLastVisit: Duration of the Last Visit (KS Statistic=0.7163)

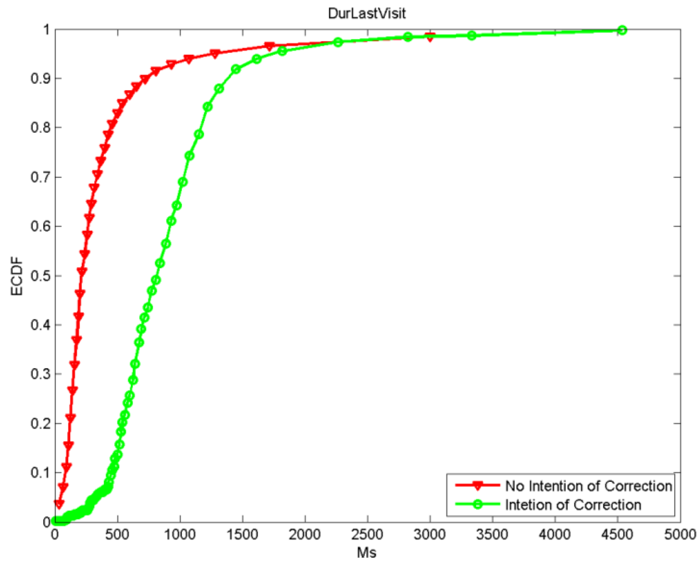


Figure A. 7: Empirical cumulative probability distributions for DurLastVisit

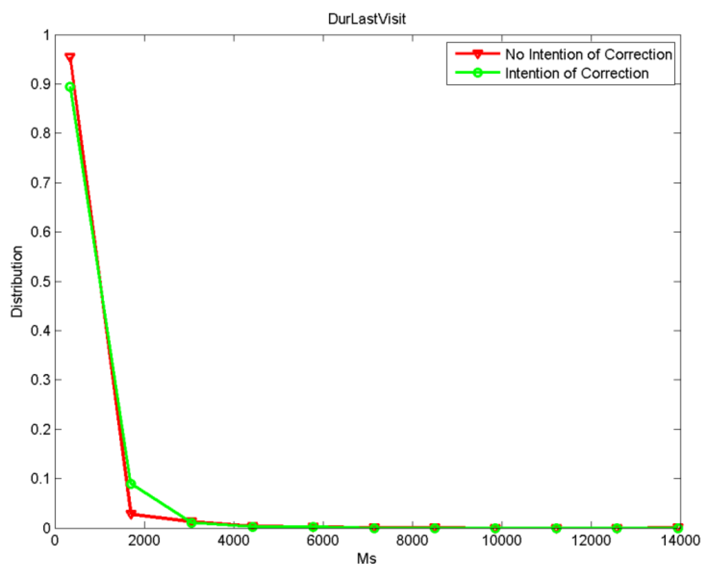


Figure A. 8: Empirical probability distributions for DurLastVisit

NofVisits:Number of Visits (KS Statistic=0.1953)

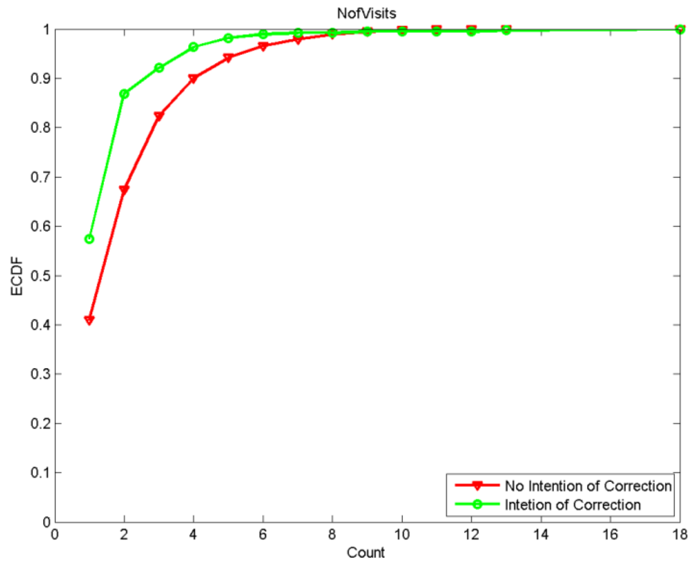


Figure A. 9: Empirical cumulative probability distributions for NofVisits

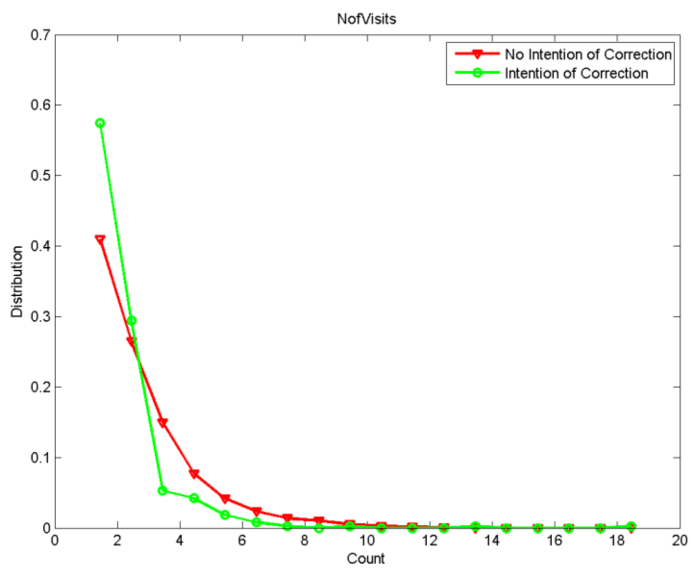


Figure A. 10: Empirical probability distributions for NofVisits

MeanDurVisit : Mean Duration of Visits (KS Statistic=0.5970)

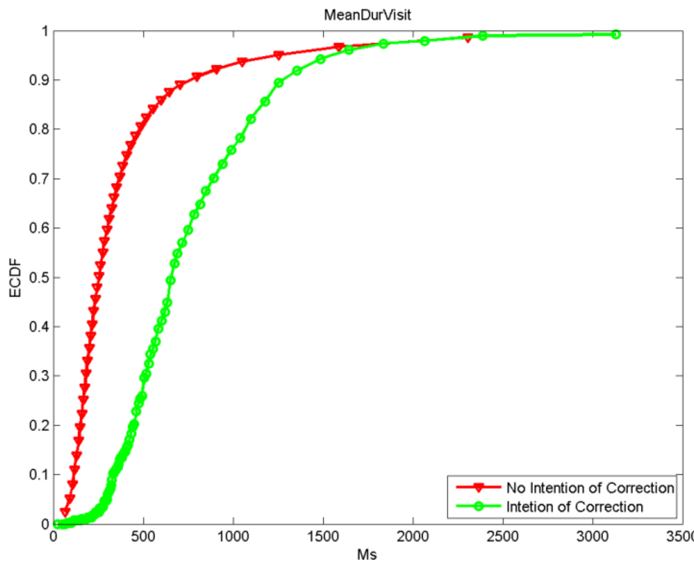


Figure A. 11: Empirical cumulative probability distributions for MeanDurVisit

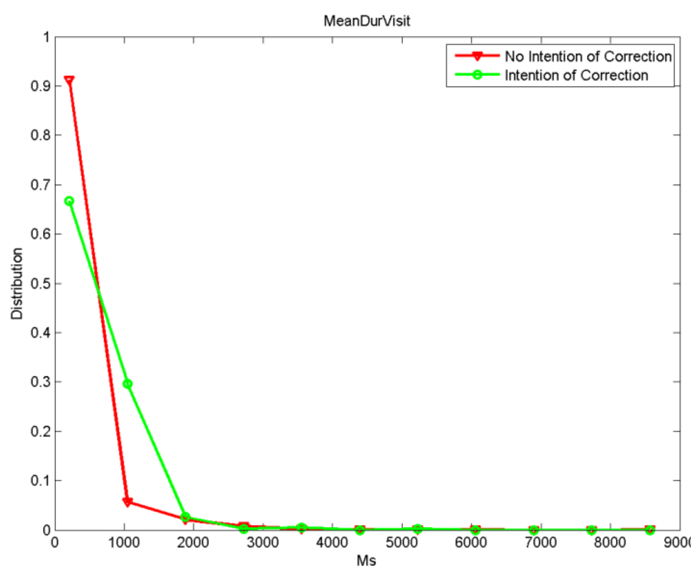


Figure A. 12: Empirical probability distributions for MeanDurVisit

FixVisit: Fixations per Visit (KS Statistic=0.1116)

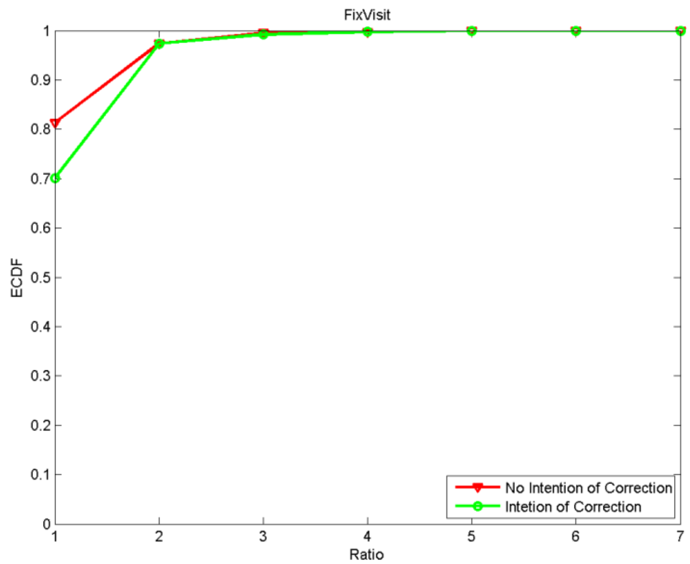


Figure A. 13: Empirical cumulative probability distributions for FixVisit

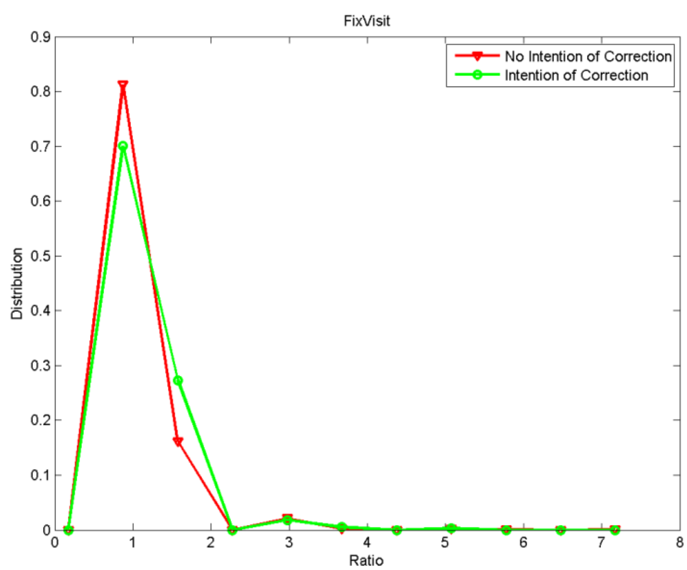


Figure A. 14: Empirical probability distributions for FixVisit

MeanSpan : Mean Span of Approaching Saccades (KS Statistic=0.1023)

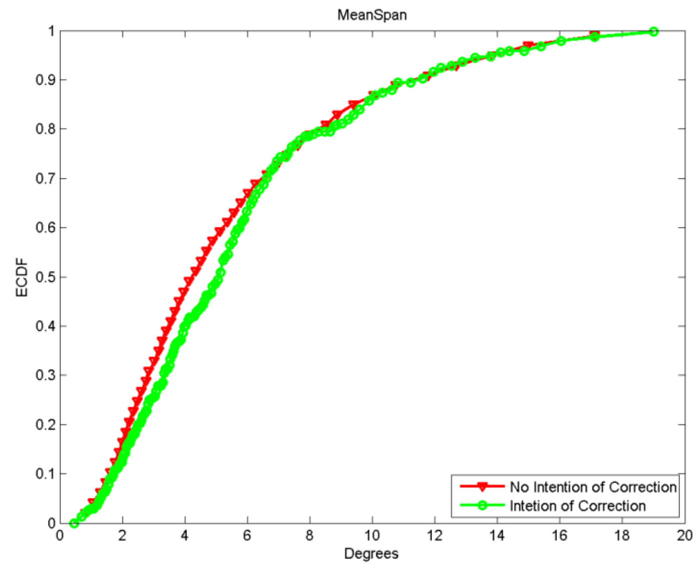


Figure A. 15: Empirical cumulative probability distributions for MeanSpan

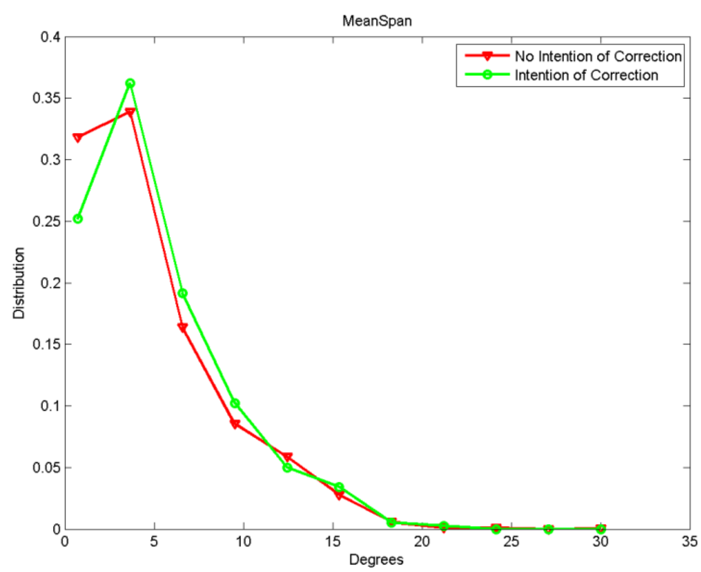


Figure A. 16: Empirical probability distributions for MeanSpan

FirstSpan : Span of the First Approaching Saccade (KS Statistic=0.1532)

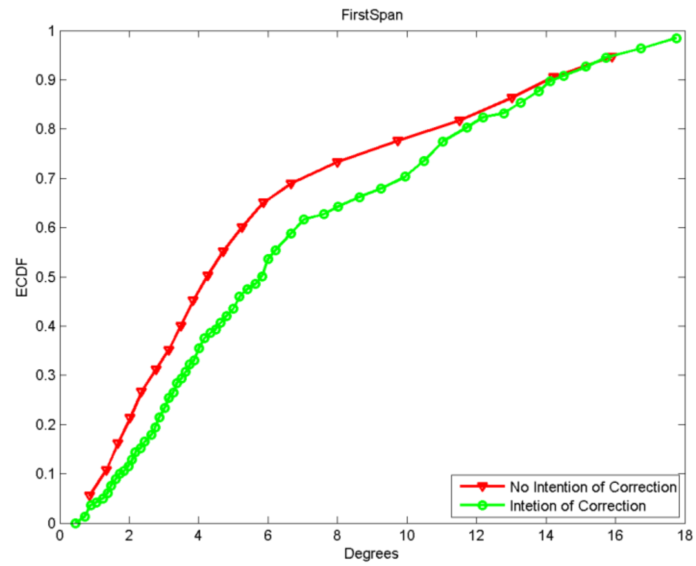


Figure A. 17: Empirical cumulative probability distributions for FirstSpan

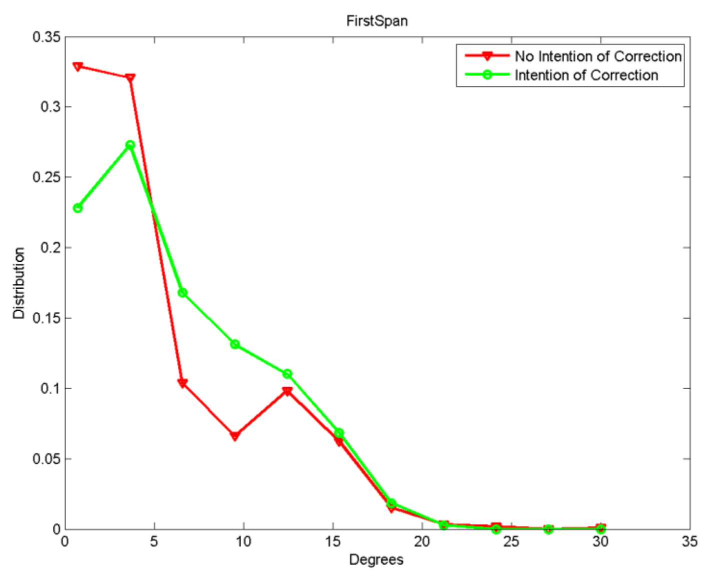


Figure A. 18: Empirical probability distributions for FirstSpan

LastSpan: Span of the Last Approaching Saccade (KS Statistic=0.0977)

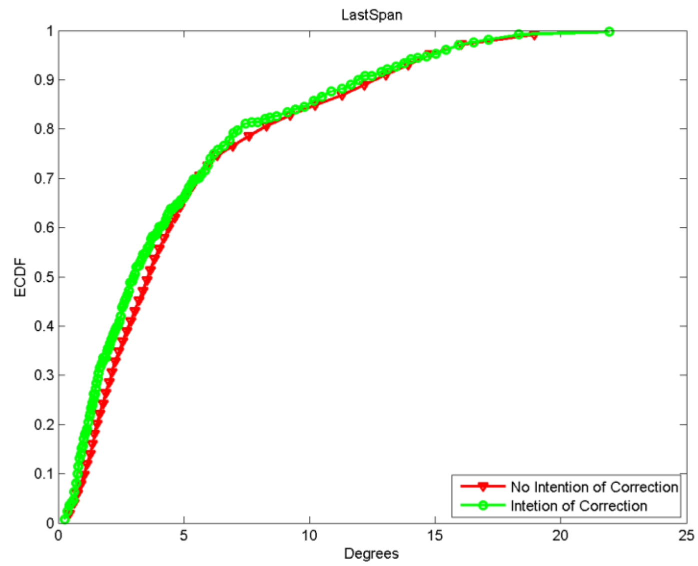


Figure A. 19: Empirical cumulative probability distributions for LastSpan

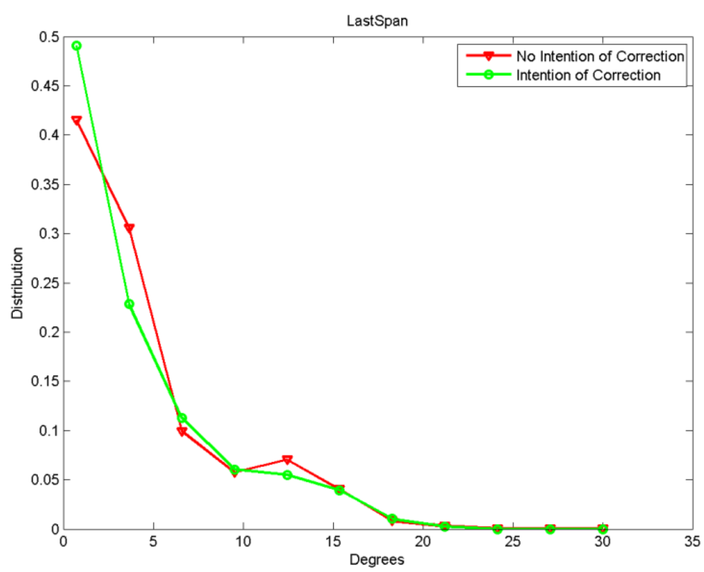


Figure A. 20: Empirical probability distributions for LastSpan

MeanSpeed : Mean Speed of Approaching Saccades (KS Statistic=0.0959)

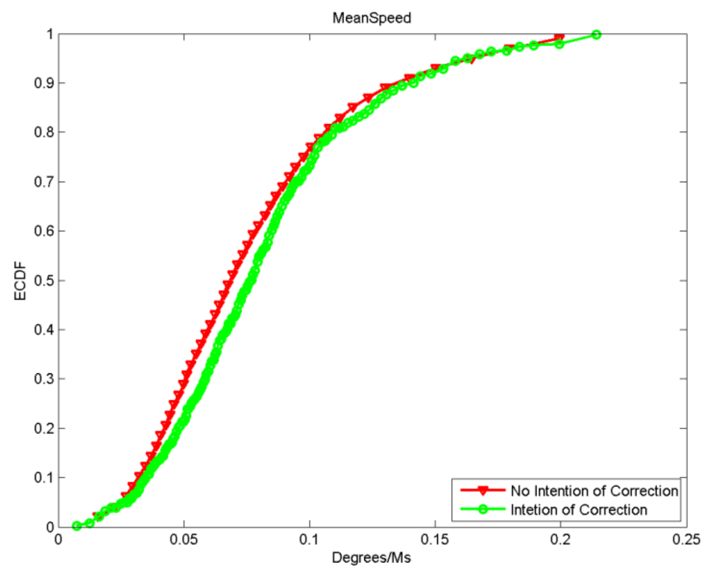


Figure A. 21: Empirical cumulative probability distributions for MeanSpeed

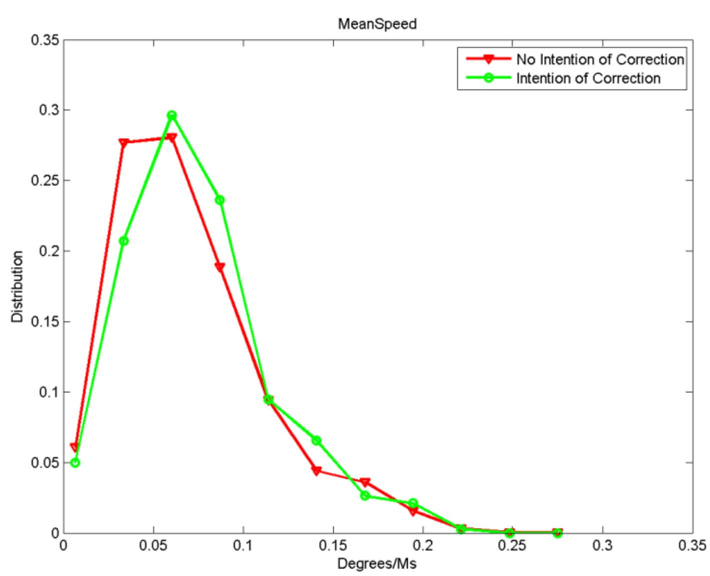


Figure A. 22: Empirical probability distributions for MeanSpeed

StdFixTime : Standard Deviation of Fixation Start Times (KS Statistic=0.3536)

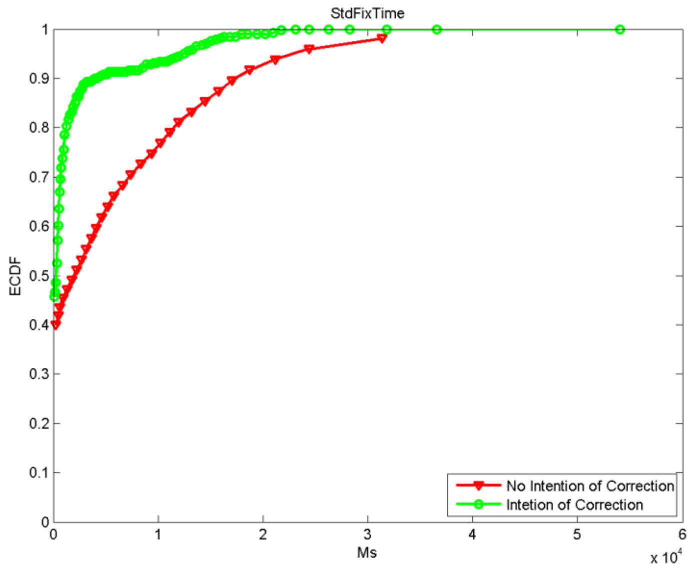


Figure A. 23: Empirical cumulative probability distributions for StdFixTime

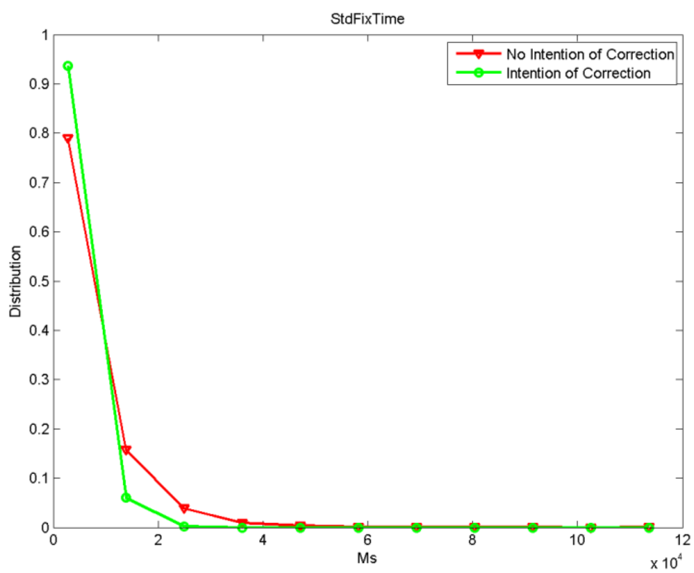


Figure A. 24: Empirical probability distributions for StdFixTime

StdVisitTime : Standard Deviation of Visit Start Times (KS Statistic=0.3517)

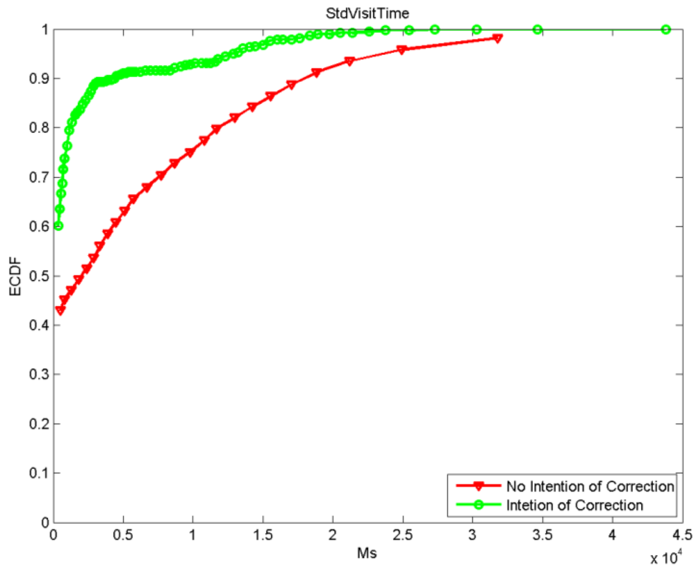


Figure A. 25: Empirical cumulative probability distributions for StdVisitTime

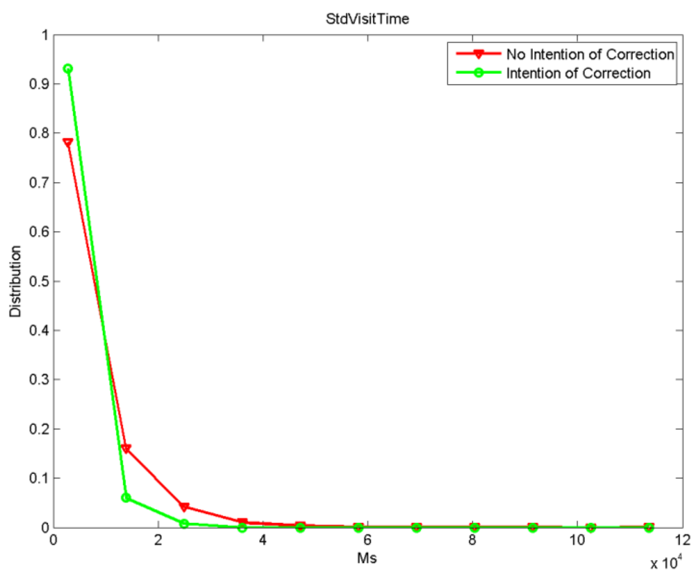


Figure A. 26: Empirical probability distributions for StdVisitTime

TimeStimFF : Time Between Stimuli and the First Fixation (KS Statistic=0.2052)

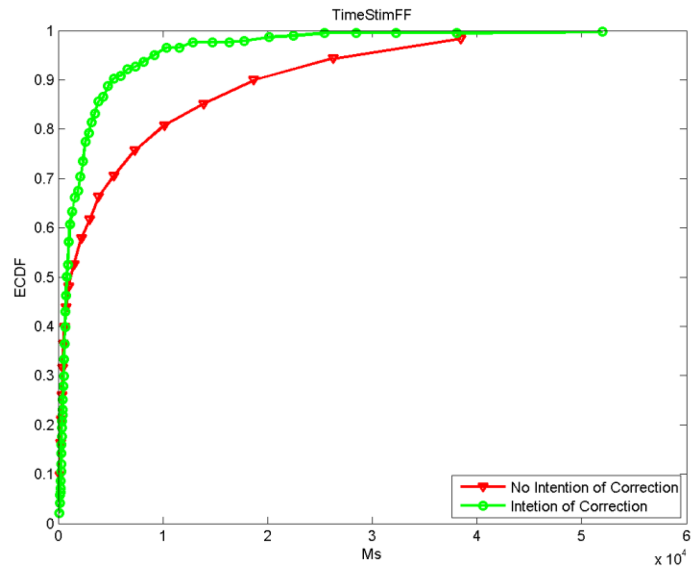


Figure A. 27: Empirical cumulative probability distributions for TimeStimFF

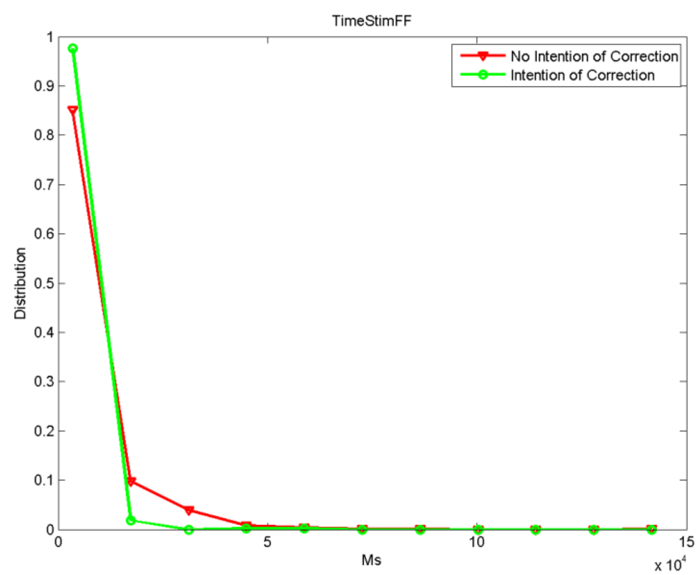


Figure A. 28: Empirical probability distributions for TimeStimFF

TimeFFCor : Time Between the First Fixation and Correction (KS Statistic=0.3403)

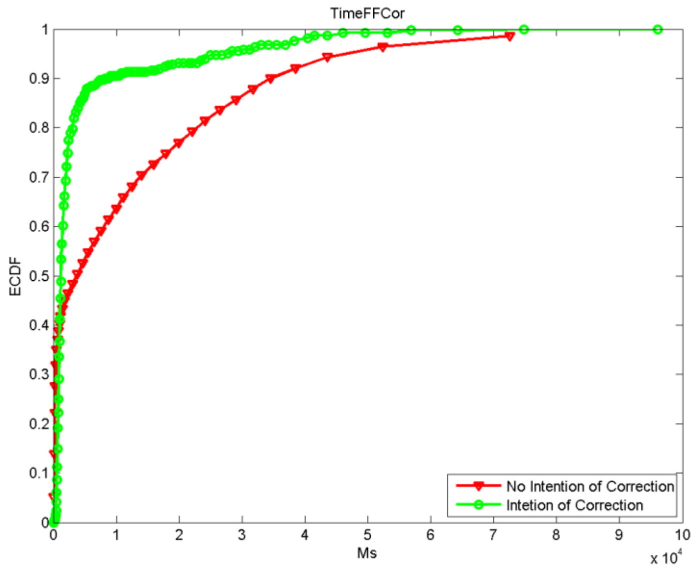


Figure A. 29: Empirical cumulative probability distributions for TimeFFCor

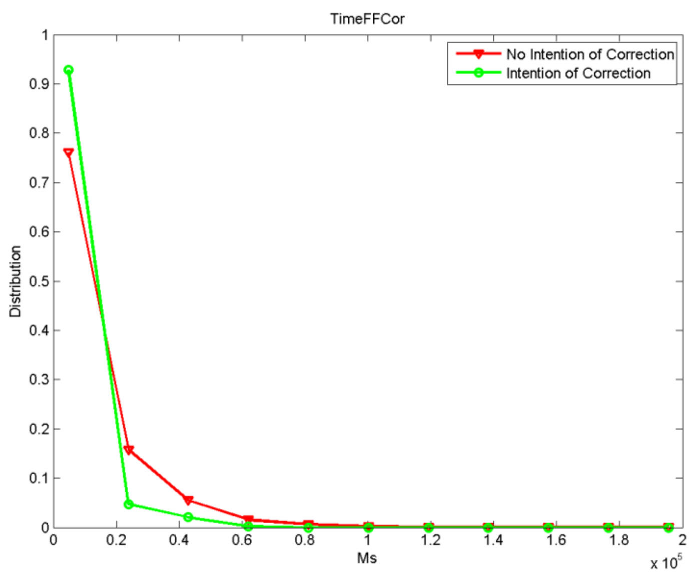


Figure A. 30: Empirical probability distributions for TimeFFCor

NofSketchElse: Number of Sketches made Elsewhere During a Fixation (KS Statistic=0.0443)

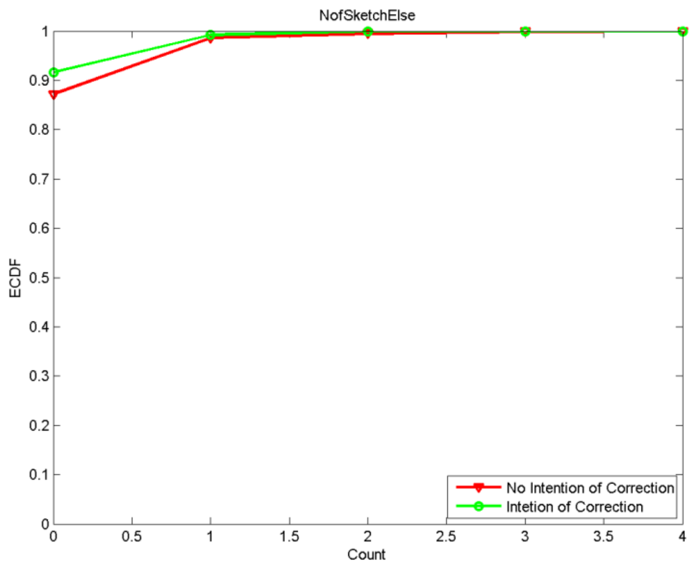


Figure A. 31: Empirical cumulative probability distributions for NofSketchElse

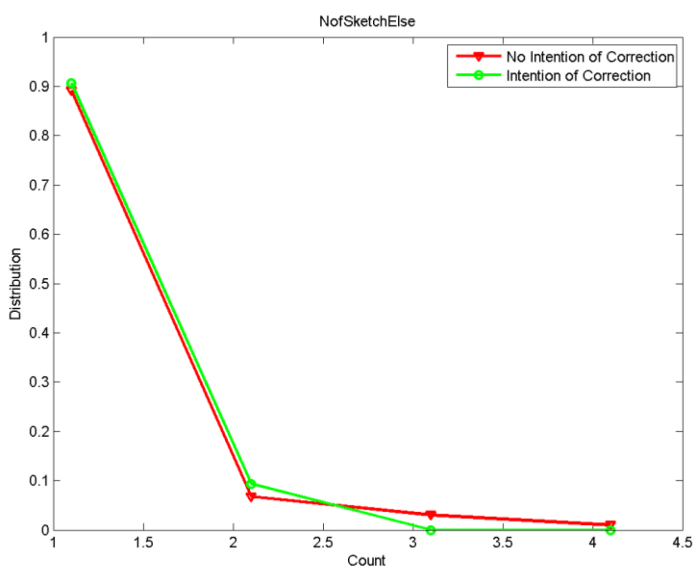


Figure A. 32: Empirical probability distributions for NofSketchElse

NofStrokesBtw : Number of Strokes Between Two Fixations (KS Statistic=0.2949)

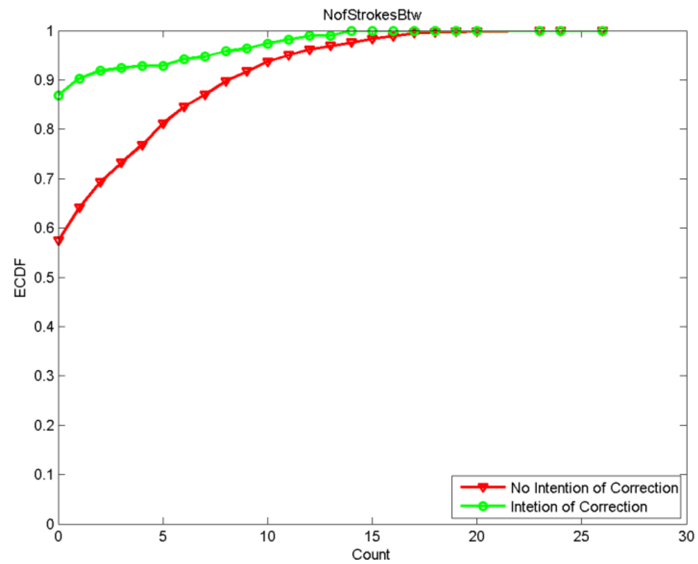


Figure A. 33: Empirical cumulative probability distributions for NofStrokesBtw

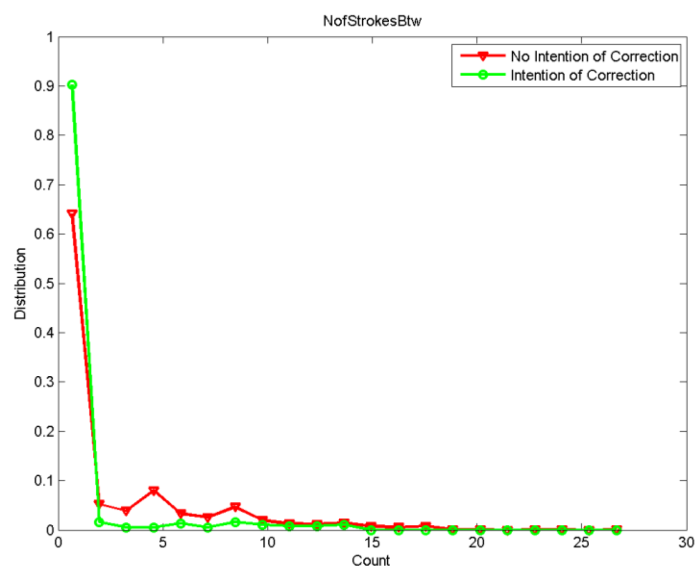


Figure A. 34: Empirical probability distributions for NofStrokesBtw

NofVisitOther : Number of Visits to Other Sketch Recognitions Between 2 Fixations (KS Statistic=0.2863)

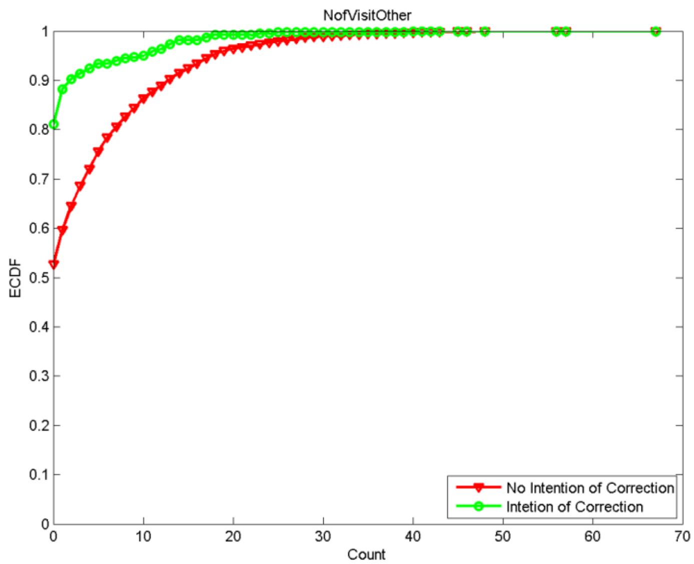


Figure A. 35: Empirical cumulative probability distributions for NofVisitOther

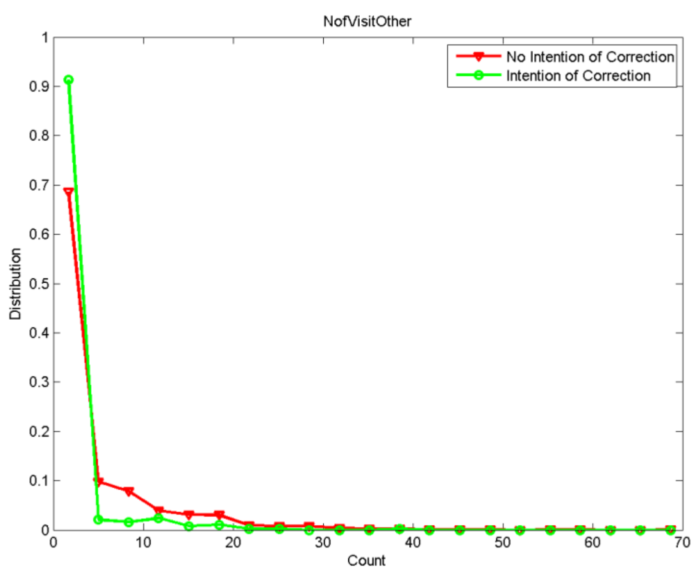


Figure A. 36: Empirical probability distributions for NofVisitOther

VisitDurOther : Duration of Visits to Other Sketch Recognitions Between 2 Fixations (KS Statistic=0.2865)

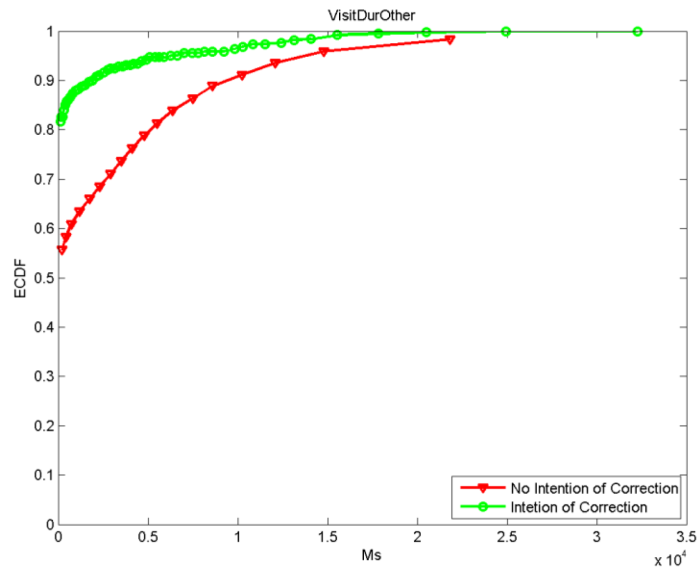


Figure A. 37: Empirical cumulative probability distributions for VisitDurOther

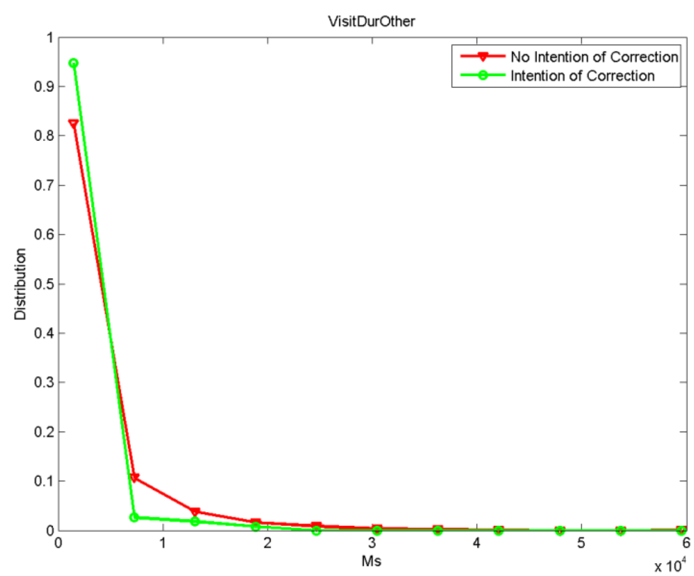


Figure A. 38: Empirical probability distributions for VisitDurOther

NofFixEmpty: Number of Fixations to Empty Space Between 2 Fixations (KS Statistic=0.3192)

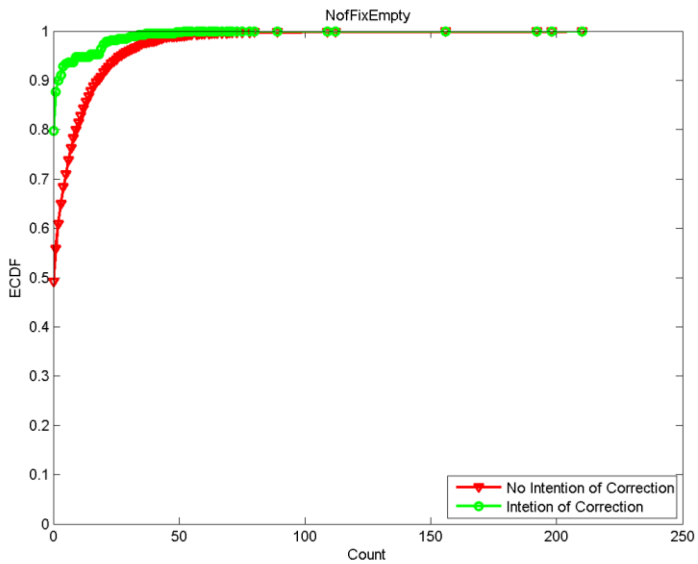


Figure A. 39: Empirical cumulative probability distributions for NofFixEmpty

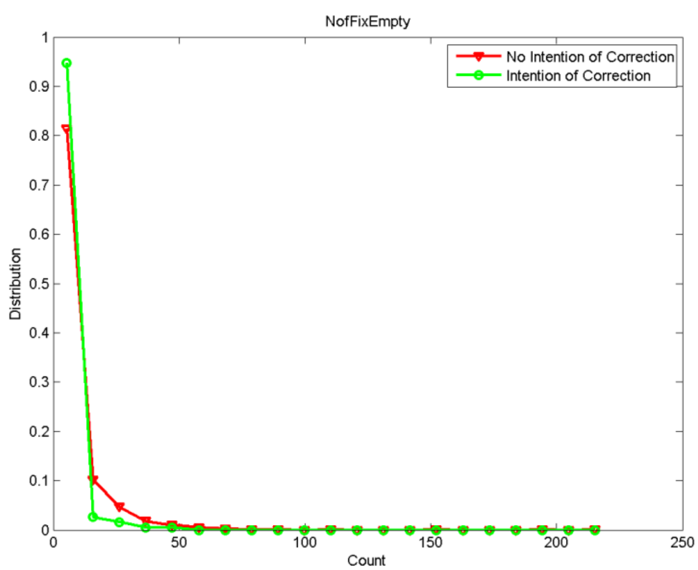


Figure A. 40: Empirical probability distributions for NofFixEmpty

FixDurEmpty: Duration of Fixations to Empty Space Between 2 Fixations (KS Statistic=0.3353)

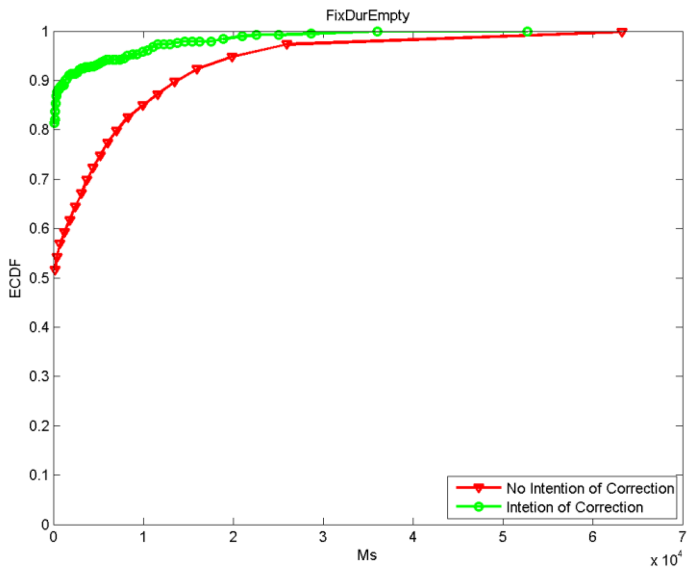


Figure A. 41: Empirical cumulative probability distributions for FixDurEmpty

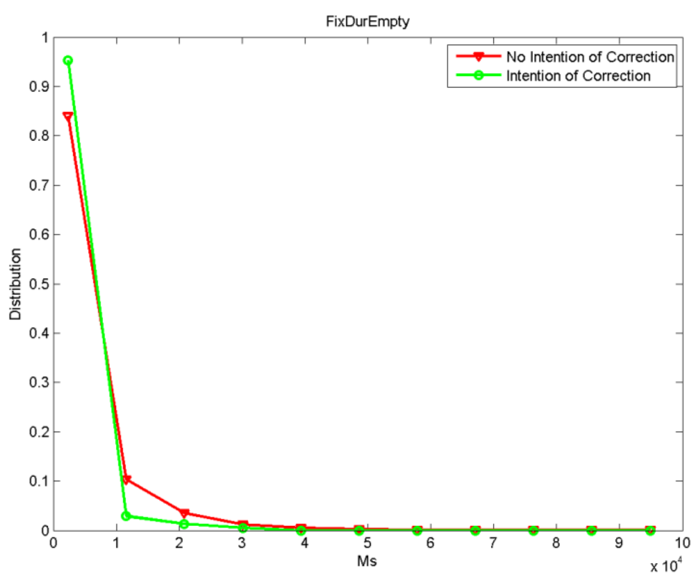


Figure A. 42: Empirical probability distributions for FixDurEmpty

BIBLIOGRAPHY

[1] S. Cheema and J. J. L. Jr, “PhysicsBook: A sketch-based interface for animating physics diagrams,” in *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces (IUI ’12)*, 2012, pp. 51–60.

[2] O. Atilola, S. Valentine, H.-H. Kim, D. Turner, E. McTigue, T. Hammond, and J. Linsey, “Mechanix: A natural sketch interface tool for teaching truss analysis and free-body diagrams,” *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, vol. 28, no. 02, pp. 169–192, May 2014.

[3] A. Blessing, M. T. Sezgin, R. Arandjelovic, and P. Robinson, “A multimodal interface for road design,” in *Proceedings of the International Conference on Intelligent User Interfaces (IUI ’09)*, 2009, pp. 1–8.

[4] A. L. Yarbus, *Eye Movements and Vision*. New York: Plenum Press, 1967.

[5] J. Henderson, “Human gaze control during real-world scene perception,” *Trends in Cognitive Sciences*, vol. 7, no. 11, pp. 498–504, Nov. 2003.

[6] B. Gesierich, A. Bruzzo, G. Ottoboni, and L. Finos, “Human gaze behaviour during action execution and observation,” *Acta Psychologica*, vol. 128, no. 2, pp. 324–330, Jun. 2008.

[7] L. Itti and C. Koch, “Computational modelling of visual attention,” *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, Mar. 2001.

[8] S. Goferman, L. Zelnik-Manor, and A. Tal, “Context-aware saliency detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.

[9] V. Navalpakkam and L. Itti, “Modeling the influence of task on attention,” *Vision Research*, vol. 45, no. 2, pp. 205–31, Jan. 2005.

[10] T. Levin and D. J. Simons, “Change blindness,” *Trends in Cognitive Sciences*, vol. 1, no. 7, pp. 261–267, 1997.

[11] D. T. Smith and T. Schenk, “Reflexive attention attenuates change blindness (but only briefly),” *Perception & Psychophysics*, vol. 70, no. 3, pp. 489–495, 2008.

[12] J. R. Brockmole and J. M. Henderson, “Prioritizing new objects for eye fixation in real-world scenes: Effects of object–scene consistency,” *Visual Cognition*, vol. 16, no. 2–3, pp. 375–390, Feb. 2008.

[13] H. Uke and M. Hayhoe, “Is attention drawn to changes in familiar scenes?,” *Visual Cognition*, vol. 16, no. 2–3, pp. 356 – 374, Feb. 2010.

[14] S. Werner and B. Thies, “Is ‘Change Blindness’ attenuated by domain-specific expertise? An expert-novices comparison of change detection in football images,” *Visual Cognition*, vol. 7, no. 1–3, pp. 163–173, Jan. 2000.

[15] I. Biederman, R. J. Mezzanotte, J. an C. Rabinowitz, C. M. Francolini, and D. Plude, “Detecting the unexpected in photointerpretation,” *Human Factors*, vol. 23, no. 2, pp. 153–164, 1981.

[16] J. M. Henderson and A. Hollingworth, “The role of fixation position in detecting scene changes across saccades,” *Psychological Science,* vol. 10, no. 5, pp. 438–443, Sep. 1999.

[17] A. Hollingworth and J. M. Henderson, “Semantic informativeness mediates the detection of changes in natural scenes,” *Visual Cognition*, vol. 7, no. 1–3, pp. 213–235, Jan. 2000.

[18] J. R. Brockmole and J. M. Henderson, “Object appearance, disappearance, and attention prioritization in real-world scenes,” *Psychonomic Bulletin & Review,* vol. 12, no. 6, pp. 1061–1067, 2005.

[19] R. J. K. Jacob, “The use of eye movements in human-computer interaction techniques: What you look at is what you get,” *ACM Transactions on Information Systems*, vol. 9, no. 2, pp. 152–169, Apr. 1991.

[20] P. Majaranta and R. Kari-jouko, “Twenty years of eye typing: Systems and design issues,” in *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA ’02)*, 2002, pp. 15–22.

[21] A. Hyrskykari, M. Paivi, A. Antti, and K.-J. Raiha, “Design issues of iDict: A gaze-assisted translation aid I-DICT,” in *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications (ETRA ’00)*, 2000, pp. 9–14.

[22] D. Rozado, J. S. Agustin, F. B. Rodriguez, and P. Varona, “Gliding and saccadic gaze gesture recognition in real time,” *ACM Transactions on Interactive Intelligent Systems*, vol. 1, no. 2, pp. 1–27, Jan. 2012.

[23] T. Toyama, D. Sonntag, A. Dengel, T. Matsuda, M. Iwamura, and K. Kise, "A mixed reality head-mounted text translation system using eye gaze input," in *Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI '14)*, 2014, pp. 329–334.

[24] H. Drewes and A. Schmidt, "Interacting with the computer using gaze gestures," in *Proceedings of the 11th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part II (INTERACT' 07)*, 2007, pp. 475–488.

[25] T. Bader, M. Vogelgesang, and E. Klaus, "Multimodal integration of natural gaze behavior for intention recognition during object manipulation," in *Proceedings of the 2009 International Conference on Multimodal Interfaces (ICMI-MLMI '09)*, 2009, pp. 199–206.

[26] A. Fathi, Y. Li, and J. M. Rehg, "Learning to recognize daily actions using gaze," in *Proceedings of the 12th European Conference on Computer Vision - Volume Part I (ECCV '12)*, 2012, pp. 314–327.

[27] C. S. Campbell, S. Jose, and P. P. Maglio, "A robust algorithm for reading detection," in *Proceedings of the 2001 Workshop on Perceptive User Interfaces (PUI '01)*, 2001, pp. 1–7.

[28] A. Bulling, D. Roggen, and G. Tröster, "What's in the eyes for context-awareness?," *Pervasive Computing*, vol. 10, no. 2, pp. 48–57, 2011.

[29] F. Courtemanche, E. Aïmeur, A. Dufresne, M. Najjar, and F. Mpondo, "Activity recognition using eye-gaze movements and traditional interactions," *Interacting with Computers*, vol. 23, no. 3, pp. 202–213, May 2011.

[30] B. Steichen, G. Carenini, and C. Conati, “User-adaptive information visualization: Using eye gaze data to infer visualization tasks and user cognitive abilities,” in *Proceedings of the 18th International Conference on Intelligent User Interfaces (IUI ’13)*, 2013, pp. 317–328.

[31] R. Bednarik, H. Vrzakova, and M. Hradis, “What do you want to do next: A novel approach for intent prediction in gaze-based interaction,” in *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA ’12)*, 2012, vol. 1, no. 212, pp. 83–90.

[32] I. Starker and R. A. Bolt, “A gaze-responsive self-disclosing display,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’90)*, 1990, no. April, pp. 3–10.

[33] P. P. Maglio, R. Barrett, S. Campbell, and T. Selker, “SUITOR: An attentive information system,” in *Proceedings of the 5th International Conference on Intelligent User Interfaces (IUI ’00)*, 2000, pp. 169–176.

[34] P. Qvarfordt and S. Zhai, “Conversing with the user based on eye-gaze patterns,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’05)*, 2005, pp. 221–230.

[35] T. Hirayama, J.-B. Dodane, H. Kawashima, and T. Matsuyama, “Estimates of user interest using timing structures between proactive content-display updates and eye movements,” *IEICE Transactions on Information and Systems*, vol. E93-D, no. 6, pp. 1470–1478, 2010.

[36] B. Brandherm, H. Prendinger, and M. Ishizuka, "Interest estimation based on dynamic bayesian networks for visual attentive presentation agents," in *Proceedings of the 9th International Conference on Multimodal Interfaces (ICMI '07)*, 2007, pp. 346–349.

[37] Y. I. Nakano and R. Ishii, "Estimating user's engagement from eye-gaze behaviors in human-agent conversations," in *Proceedings of the 15th International Conference on Intelligent User Interfaces (IUI '10)*, 2010, pp. 139–148.

[38] K. Puolamaki, J. Salojarvi, E. Savia, J. Simola, and S. Kaski, "Combining eye movements and collaborative filtering for proactive information retrieval," in *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*, 2005, pp. 146–153.

[39] A. Klami, C. Saunders, T. E. Campos, and S. Kaski, "Can relevance of images be inferred from eye movements ?," in *ACM International Conference on Multimedia Information Retrieval (MIR'08)*, 2008, pp. 134–140.

[40] G. T. Papadopoulos, K. C. Apostolakis, and P. Daras, "Gaze-based relevance feedback for realizing region-based image retrieval," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 440–454, 2014.

[41] P. Wais, A. Wolin, and C. Alvarado, "Designing a sketch recognition front-end : User perception of interface elements," in *Proceedings of the 4th Eurographics Workshop on Sketch-based Interfaces and Modeling (SBIM '07)*, 2007, pp. 99–106.

[42] R. C. Davis, T. S. Saponas, M. Shilman, and J. A. Landay, "SketchWizard: Wizard of Oz prototyping of pen-based user interfaces," in *Proceedings of the 20th Annual*

ACM Symposium on User Interface Software and Technology (UIST ’07), 2007, pp. 119–128.

[43] D. D. Salvucci and J. H. Goldberg, “Identifying fixations and saccades in eye-tracking protocols,” in *Proceedings of the Symposium on Eye tracking Research & Applications (ETRA ’00)*, 2000, pp. 71–78.

[44] J. Biing-Hwang and L. R. Rabiner, “The segmental k-means algorithm for estimating parameters of hidden Markov models,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 9, pp. 1639–1641, 1990.

[45] A. Hollingworth, “Failures of retrieval and comparison constrain change detection in natural scenes.,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 29, no. 2, pp. 388–403, 2003.

[46] L. Kozma, A. Klami, and S. Kaski, “GaZIR: Gaze-based zooming interface for image retrieval,” in *Proceedings of the 2009 International Conference on Multimodal Interfaces (ICMI-MLMI ’09)*, 2009, pp. 305–312.

[47] C. L. Wood and M. M. Altavela, “Large-sample results for Kolmogorov-Smirnov statistics for discrete distributions,” *Biometrika*, vol. 65, no. 1, pp. 235–239, 1978.

[48] C. Chang and C. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–39, 2001.