MEDIATING DISRUPTION IN HUMAN-COMPUTER INTERACTION FROM IMPLICIT METRICS OF ATTENTION

by

Ernesto Arroyo Acosta

B.S., Universidad Autónoma de Baja California, 1999
M.S., Massachusetts Institute of Technology, 2002

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences
at the
Massachusetts Institute of Technology
September 2007

AUTHOR: **ERNESTO ARROYO ACOSTA**
Program in Media Arts and Sciences
August 20, 2007

CERTIFIED BY: **TED SELKER**
Associate Professor
Thesis Supervisor

ACCEPTED BY: **DEB ROY**
Chairperson
Departmental Committee on Graduate Students

MEDIATING DISRUPTION IN HUMAN-COMPUTER INTERACTION FROM IMPLICIT METRICS OF ATTENTION

by

Ernesto Arroyo Acosta

B.S., Universidad Autónoma de Baja California, 1999
M.S., Massachusetts Institute of Technology, 2002

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences
at the
Massachusetts Institute of Technology
September 2007

THESIS READER: **ROBERT JACOB**
Associate Professor
DEPT. OF COMPUTER SCIENCE
TUFTS UNIVERSITY

MEDIATING DISRUPTION IN HUMAN-COMPUTER INTERACTION FROM IMPLICIT METRICS OF ATTENTION

by

Ernesto Arroyo Acosta

B.S., Universidad Autónoma de Baja California, 1999
M.S., Massachusetts Institute of Technology, 2002

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning

In partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences
at the
Massachusetts Institute of Technology
September 2007

THESIS READER: **ANDREW B. LIPPMAN**
Associate Professor
PROGRAM IN MEDIA ARTS AND SCIENCES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MEDIATING DISRUPTION IN HUMAN-COMPUTER INTERACTION FROM IMPLICIT METRICS OF ATTENTION

by

Ernesto Arroyo Acosta

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning
on August 20, 2007, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Media Arts and Sciences

Abstract

Multitasking environments cause people to be interrupted constantly, often disrupting their ongoing activities and impeding reaching their goals. This thesis presents a disruption reducing approach designed to support the user's goals and optimize productivity that is based on a model of the user's receptivity to an interruption. The model uses knowledge of the interruption content, context and priority of the task(s) in progress, user actions and goal-related concepts to mediate interruptions.

The disruption management model is distinct from previous work by the addition of implicit sensors that deduce the interruption content and user context to help determine when an interruption will disrupt an ongoing activity. Domain-independent implicit sensors include mouse and keyboard behaviors, and goal-related concepts extracted from the user documents. The model also identifies the contextual relationship between interruptions and user goals as an important factor in how interruptions are controlled. The degree to which interruptions are related to the user goal determines how those interruptions will be received.

We tested and evolved the model in various cases and showed significant improvement in both productivity and satisfaction. A disruption manager application controls interruptions on common desktop computing activities, such as web browsing and instant messaging. The disruption manager demonstrates that mediating interruptions by supporting the user goals can improve performance and overall productivity. Our evaluation shows an improvement in success of over 25% across prioritization conditions for real life computing environments.

Goal priority and interruption relevance play an important role in the interruption decision process and several experiments these factors on people's reactions and availability to interruptions, and overall performance. These experiments demonstrate that people recognize the potential benefits of being interrupted and adjust their susceptibility to interruptions during highly prioritized tasks.

The outcome of this research includes a usable model that can be extended to tasks as diverse as driving an automobile and performing computer tasks. This thesis supports mediating technologies that will recognize the value of communication and control interruptions so that people are able to maintain concentration amidst their increasingly busy lifestyles.

Thesis Supervisor: **Ted Selker**
Title: Associate Professor, MIT Program in Media Arts and Sciences

TABLE OF CONTENTS

LIST OF FIGURES

C H A P T E R

1

INTRODUCTION

1.1 Human Interruption

In Human Computer interaction, *an interruption can be defined as an unanticipated request for task switching while multitasking* (Bailey 2000; McFarlane 1999; Latorella 1998; Gillie and Broadbent 1989). Interruptions typically "request immediate attention" and "insist on action" (Covey 1989). Thus, a person, an object, or an event can create an interruption—the timing of which is beyond the user's control. Furthermore, an interruption diverts people’s attention from an ongoing activity and compels people to turn their attention towards an interrupting task.

Coordinating interruptions involves one or more person’s modes of activity: cognition, perception, or physical action. Interruption coordination is the method by which a person shifts their focus of consciousness from one processing stream to another. *Disruption is the negative effects on a primary task from interruptions requiring transition and reallocation of attention focus from one activity to another.*

It is important to differentiate between distractions, interruptions and disruption to allow for a framework for understanding human disruption. Distractions and interruptions are similar in that they can both occur while the user performs a primary activity. “Distraction conflict theory” describes a research stream investigating the influence of distractions and specifies that distractions are detected by sensory channels different from those used by the primary activity. Distractions can be ignored or processed at the same time as the primary activity (Cohen 1980; Groff, Baron and Moore 1983). As opposed to distractions, interruptions share the same sensory channel as the primary activity and encompass an interrupting task that should be performed. People cannot choose to ignore interruption cues, resulting in capacity and structural interference, disrupting of the ongoing activity and often negatively affecting human performance. Capacity interference occurs when the number of incoming cues is greater than people can process. In addition, structural interference occurs when people must attend to two inputs that require the same physiological mechanisms (Kahneman 1973), such as attending to two different visual signals—a computer screen and a public ambient display.

Through the remainder of this thesis, we will use the term disruption to accommodate a situation in which people have accepted an unanticipated request for task switching from an interrupting task, causing negative effects on the ongoing activity.

Interruptions are an everyday and normal part of human behavior. People frequently interrupt communication dialogue, such as an unanticipated request for topic switching while having a conversation. Interruptions are also common in computing environments; users can commission another person or a computerized intelligent agent to perform tasks on their behalf to avoid cognitive overload and successfully perform multiple tasks. In this context, interruptions might be seen as the side effect of delegating tasks to intelligent agents (Maes 1994, Selker 1994).

This thesis focuses on interruptions presented to users by computer technologies as the result of a request for synchronous or asynchronous communication. We have defined this type of interruptions as user-initiated and technology-mediated.

The term technology-mediated refers to interruptions originating from a computer application as a request from another user, such as email, or IM notifications. The term user-initiated excludes interruptions originating from a computer application due to an internal error, or status change. Our definition also excludes interruptions originating from the external environment such as a telephone call or knock at the door. Interruptions on computer settings provide a complex set of situations that might be prone to productivity loss.

Careful design of interrupting notifications might help reduce disruption effects on people's performance on ongoing activities. How efficiently and effectively interruptions are handled by users might depend on characteristics of the notification itself and characteristics of the ongoing activity. However, people's reactions to interruptions and perceived disruption are principally affected by goal-oriented strategies users adopt to evaluate incoming interruptions in order to accomplished their goals (Latorella 1998; McCrickard and Chewar 2003). For example, a person who works at an office is more likely to take an incoming phone call from a co-worker while at work than when he is on his way home. It is common for people to juggle several competing goals at one given time, while their priorities might change depending on various factors. This is exemplified in a diary study of office work, which reported frequent and deliberate task-switching activities (Czerwinski, Horvitz, and Wilhite 2004). Similarly, residential interviews and self-reports revealed that willingness to handle interruptions varies across individuals with current location, as well as with current activity (Nagel, Hudson, and Abowd 2004).

1.2 Proposed Research

This work is aimed at understanding the underlying factors that influence people when dealing with interruptions. An interruption model sets the framework for exploring different aspects of human disruption in desktop interaction. This thesis focuses on evaluating the requirements from interruptions regarding their context, and their relationships to the user goals and tasks. The work takes the approach that goal concepts, together with task context serve as important factors for reasoning and predicting disruption.

This thesis explores the use of implicit sensors to provide inferences about the scope of people's goals and tasks. Such sensors derive information regarding the user activity from normal behavior without the addition of new sensors. The inferences generated from virtual sensors use domain-independent implicit metrics of interaction. For demonstration purposes, the work explores the inferences that can be drawn from low level mouse metrics, such as predicting user interest, and ongoing activity.

This thesis argues that goal and task priority play an important role in the interruption decision process, and evaluates how changes in task priority affect people's reactions to interruptions. The thesis also evaluates the effects from interruption relevancy to the users' goal and task priority on perceived disruption. Finally, this work presents and evaluates a disruption manager based on our findings regarding disruption. The manager goal is to reduce disruption and increase overall user satisfaction. The manager outlines the factors needed to mediate disruption in computing activities under different contexts. A disruption manager would mediate instant message interruptions as people navigate the web. Our manager is based on research investigating people's reactions to interruptions and factors involved in the interruption process, such as interruption relevancy and task priority. The manager supports monitoring ongoing behaviors using implicit metrics to control possible disruptive outcomes given the user and system state (mouse and keyboard behaviors, interruption type: task request or notification, concepts surrounding the user's goals, interruption relevancy, and concept priority). The disruption manager selects appropriate interruption timing and whether or not instant message interruptions should be presented to the user based on concept relevancy, user engagement and task level.

1.3 Motivation

Multitasking is useful in keeping up with current demanding environments; however, it also introduces the side effect of being interrupted constantly. Unfortunately, people have cognitive limitations that make them susceptible to errors when interrupted. In general, current computer interfaces and environments are becoming more and more complex. Additionally, there are increasing number of tasks and issues which computer users have to keep track of (Maes 1994).

The use of interruptions is very important in the design of human-computer interfaces. Although interruptions sometimes might be perceived as useful (such as when introducing breaks that promote productivity), interruptions are perceived as disruptive. Simply put, people perform slower on interrupted tasks than on uninterrupted tasks.

Advances in computer technologies have enabled the creation of systems that allow people to perform multiple activities at the same time. Users delegate more and more control to automating applications such as interface agents, softbots, and peripheral information displays. Automated applications generate interruptions, as they must accomplish some of the following functions:

gain user attention periodically in order to receive additional guidance from the user,

provide feedback regarding decisions made in the user's behalf (Horvitz 1999),

keep the user aware of peripheral information (Bailey, et al 2000; Maglio and Campbell 2000).

These multitasking computing environments require a user to understand their context quickly in order to make decisions and might cause high cognitive loads. Work environments are more and more complex, with an increasing number of applications and an increasing number tasks competing for people's attention; putting people at risk of being interrupted constantly. These tasks depend on human memory and attention in order to be completed successfully; however, humans have limited resources and examine the world through a limited spot light of attention (Horvitz, et al 1999; MacCrickard and Chewar 2003). Current technologies allow for an increasing amount of information available to users at all times. This information is often distracting and its effects are not well understood.

1.3.1 Increasing Workload and Interruptions

Interruptions are common to today's multitasking computing user interface experience. This kind of multitasking environment is useful and might seem natural, however it also introduces the side effect of causing people to be interrupted constantly. Unfortunately, people have cognitive limitations that make them susceptible to errors when interrupted. For instance, some researchers (McFarlane 1999; Czerwinski 2000-A) have examined interruptions by looking at when to interrupt users in a multitasking environment.

Various effects of interruptions have previously been studied in both psychology and human-machine interaction. Some of these studies have inconsistent conclusions and further investigation into the effects of an interruption on a user's task performance is necessary. Zijlstra, measured the effects of interruption frequency and complexity on a user's emotional state and task performance. The authors found that interrupting users during a document-editing task caused them to complete the task faster than when performing the same task without interruption (Zijlstra, et al 1999). The authors suggest that participants developed strategies to deal with the interruptions, thus compensating performance decline. They also found that interruptions have a negative impact on emotion and well-being. In contrast to Zijlstra, Kreifeldt and McCarthy found that interrupting a user while performing another type of task, like a series of calculator-based problems, caused that user to complete those tasks slower than when performing without interruptions (Kreifeldt and McCarthy 1981). These experiments illustrate differences in the way people deal with interruptions and suggest further investigation on the factors affecting the decision process is necessary.

1.3.2 Adaptive Interfaces

In one of the first works on adaptive interfaces, Greenberg and Witten proved the viability of providing adaptive user interfaces. The authors made a study on telephone usage patterns and found that in normal usage many numbers had been dialed previously and built a system that allowed access to a directory of telephone numbers through a hierarchy of menus. They presented items at a level in the hierarchy corresponding to the chances of being selected. The higher the availability associated with an item, the earlier it could be available for selection. Their system identified differences among users and the type of adaptation used was based on probability (Greenberg and Witten 1985). Even though more recent work has explored the area of adaptive user interfaces (Münch, et al 1997 and Ramstein, et al 1996, Selker 1994), they do not consider adapting the output modality itself.

This thesis shows that it is possible to dynamically control interruptions based on implicit metrics for user's performance and disruption. Thus, it is possible to maximize the effectiveness of an interruption through proper interruption mediation and notification configuration.

Consider a student working on his final project paper about voting interfaces. He spends some time looking for references online about electoral systems, writes a few paragraphs and then decides to search for information about electoral failures in history. He finds an interesting article online and decides to read it thoroughly as indicated by his mouse movement around the webpage. As he is reading the article, an automated web-bot sends him an instant message about exciting new features of his IM client. A disruption manager determines that the content of the message is not relevant to the concepts in previously opened and current documents, and neither relevant to the content of previously accessed web pages. Thus, the web-bot message is delayed until the user is finished reading. During this time, a workmate sends the student an instant message about some electoral voting references on the Caltech-MIT website that he just found and indicating him to look at them. A disruption mediator determines that the message is related to the concepts in the student's activities and allows the instant message to go trough. The student interrupts his activity, disregards the web-bot message, reads the new references and finishes the introduction section of his paper, completing one of his or her goals continuing onto the next task.

1.4 Overview of the Dissertation

This thesis starts with a brief introduction of the topic area in which we point out that communication without disruption management is becoming extremely complex with a cacophony of interruptions that distract us from being productive. Certainly human interruption has always been an issue in human communication, and a section describes related interruptions research in the area of human interaction.

The proposed research goals located in the next section, which describes our goal of making a disruption manager based on a model of interruptions, experimentation and tools, along with an AI architecture for solving the communication problems, called disruption manager. The motivation and vision section lays out these ideals in more detail:

The value of technology for modeling aspects of interruption

The value of recognizing different behaviors in humans

The value for creating an understanding of when to interrupt

The value of pacing communication between a person and whomever's interacting with them.

This we believe will change the very nature of communication and of notification systems. We believe that all communication will benefit from mediation with disruption manager style alerts and that within the next ten years we expect that computers will efficiently manage undesired distractions.

1.5 Document Overview

This chapter has defined interruptions, described the topic area, and motivated the need for interruption-mediated computer interfaces.

Chapter 2 describes existing empirical HCI work related to interruptions, and a description of systems designed to take interruptions into consideration.

Chapter 3 presents our own previous work; including a brief overview of systems and exploratory experiments. This chapter describes how our exploratory work motivated and provided insight for the development of chapter 4.

Chapter 4 presents a disruption model that guides the development of a disruption management framework described in chapter 5.

Chapter 6 presents experiments evaluating the disruption process, these experiments validate the disruption model described in Chapter 4.

Chapter 7 validates low level implicit metrics and demonstrates several systems using mouse tracks as predictors.

Chapter 8 describes the design and implementation of a disruption manager that exemplifies our approach and includes behaviors from our empirical experimentation; evaluation is also discussed.

C H A P T E R

2

RELATED WORK

2.1 Introduction

This chapter presents research that is relevant to this thesis and informs a theory of disruption mediation. The chapter specifically focuses on Human Computer Interaction work related to interruptions, such as notification systems, mediating agents, and computer interfaces.

2.2 Human Interruption

Previous research has focused on experiments using abstract tasks under controlled environments to investigate how interruptions affect human behavior. Zeigarnik for example, described the relation between interruptions and selective memory (Zeigarnik 1927). His work described an interesting psychological phenomenon "the Zeigarnik phenomenon" that people can recall details of interrupted tasks better than those of uninterrupted tasks (Van Bergen 1968). Researchers have since documented other effects of interruption.

Hess and Detweiler found that if an interruption imposes a high memory load on the user, it is harmful to the primary task. They showed that interruptions that were similar to an ongoing computer task are quite disruptive over the first two of three sessions, but are significantly less disruptive by the third session. In addition, they found that, if participants are allowed to train on the primary task without interruptions for two sessions, a third session with interruptions is still significantly harmful to performance, despite the user being highly trained (Hess and Detweiler 1994).

In one of the most relevant work about interruptions, Gillie and Broadbent examined the features of interruptions that make them more or less disruptive to an ongoing computer task. They found that the length of an interruption does not affect how disruptive interruptions are perceived and that interruptions with similar content as the primary task can be quite disruptive even if extremely short. The authors concluded that subjects make use of some form of non-articulatory memory that is affected by processing similar material and by imposing high demands on working memory (Gillie and Broadbent 1989).

Conversely, to the work presented by Gillie and Broadbent, recent work from Czerwinski found that interruptions that were relevant to ongoing tasks can also be less disruptive than those that were irrelevant. Czerwinski designed an experiment to measure disruption caused by instant messaging interruptions. The researchers used the times required for the user to move from the primary task to the notification, the time to read notifications, and the time to return to the primary task, as measures of disruption. A significant difference in these works is the interrupting task relevancy to the ongoing activity. In one case, the interrupting task interferes with the primary task modality, whereas in Czerwinski's case, the interrupting task is directly related to the primary task, thus, the interruption is useful to the primary task. Czerwinski also found that the costs of disruption depended on the type of the ongoing task or subtask being performed (Czerwinski 2000-B).

More recently, Brian Bailey, et al, described an experiment measuring the effects interruptions on users' task performance, annoyance, and anxiety in the user interface. The researchers corroborated existing findings that users perform interrupted task more slowly than non-interrupted tasks, and that the level of annoyance experienced by a user depends on both the category of the primary task being performed and the time at which a peripheral task is displayed. They found that users experience a greater increase in anxiety when a peripheral task interrupts the primary task than when it does not. They also found that users perceive an interrupted task to be more difficult to complete than a non-interrupted task (Bailey, et al 2000-A). Their work indicates that interruptions have an effect on perceived disruption and perceived task difficulty. These subjective factors should be taken into consideration in designing mediating systems.

Work in the area of disruptions has shifted from controlled experiments that describe interruptions towards the use of sensor-based models of user activity to predict user disruption. This type of approach is often dependent on the type of sensors used and does no scale to other environments outside the test bed used to gather data. Horvitz developed a utility for mediating the flow of potentially distracting alerts and communications, such as email, to computer users by asking people the amount of money they would pay not to be interrupted. Horvitz took the perspective that human attention is the most valuable and scarcest resource available in human computer interaction, and that alerts are valuable information, but at a cost of interruption. Based on this assumption, they presented Bayesian models that balance the context sensitive cost of interruption (Horvitz, et al 1999)

More recently, Forgarty examined task engagement in sensor-based statistical models of human interruptibility. He examined how programmers respond to interruptions while they programming. He then, developed a statistical model of their interruptibility based on low-level input events (key strokes, compiler calls, etc) in a development environment (Fogarty 2005). They used a top-down approach by first studying ongoing software programming activity and identifying what sensors to use (from a large array of sensors), later developing and validating sensors on the pre-defined activity.

Iqbal and bailey investigated the characteristics of task structure to predict the cost of interruption. They found that interrupting at boundaries with lower workload results in significantly lower disruption than at boundaries with higher workload. They used specialized hardware to measure workload using physiological metrics, such as pupil dilatation. They also used statistical methods (linear regression), but included task structure to objectively predict interruption as measured by resumption lag. However, their approach is limited to tasks with prescribed execution sequences (Iqbal and Bailey 2005).

Gievska, also used Bayesian Belief Networks as a decision-support aid for selecting the best timing to mediate interruptions. They presented an exploratory user study examining the qualitative gains of mediating human interruptions by using an interruption mediator (Gievska 2005). They used experimental tasks similar to military computer games and simulations. Although not objectively validated, their subjective metrics of disruption suggested that their mediator was effective in decreasing some disruptive effects. However, since user independent sensors do not take into account user context, this resulted in subjects expressing lack of satisfaction regarding how interruptions were processed.

2.3 Notifications

The area of notifications has been studied extensive (Chewar 2004, McCrickard 2003). Several studies in this area have shown that the nature of the presentation of notifications influences performance on the primary computing task. Maglio and Campbell demonstrated that continuously scrolling displays were more distracting than discrete displays to ongoing word editing tasks. They found that all notification styles reduced word-editing performance in comparison to a no-notification condition (Maglio and Campbell 2000). Ware, et al reported an experiment designed to test the use of simple linear motion as an attention-getting device for computer displays. A primary task required the transcription of a document typed into a word processor and a secondary task involved detecting and responding to a moving icon signal. The icon was a rectangular bar that grew and shrank vertically in an oscillatory fashion. Both the amplitude and velocity of the icon's motion were varied and response time recorded. The results showed that there was a direct relationship between the velocity of the moving icon and rapidity of the response, but no effect was found for amplitude. Observed response speeds appeared to indicate that simple motion was an effective attention-getting device for events in the periphery of the visual field (Ware, et al 1992). Selective attention is highly responsible for adjusting the way how people react to notifications. During the initial stages of web browsing, web banner were extremely obtrusive and hard to ignore; however, now they hardly have any impact on people's web browsing behaviors. This thesis makes use of existing notification design, as notification design is an area of work constantly changing with people adjusting their reactions to these systems over time. However, notification design is not the main aim of this work.

2.4 Context-Aware Computing

The context-aware computing group has focused on identifying scenarios where a system has the potential to anticipate the user actions or act in liu of the user. The context-aware approach focuses on the potential user benefits, as well as identifying sensors that would provide sufficient information to support intelligence in order to fulfill the scenario and its goals (Lieberman and Selker 2000). This approach is complemented by constant evaluations that inform the design and the assumptions for a context aware system. Several examples from the context aware-field show some of the benefits by this approach. A smart door acting as an office assistant is able to control visitor interruptions based on social cues and information about the office schedule. This smart door demonstrates the value of recognizing and respecting the threshold as the most important social demarcation (Yan and Selker 2000).

Even if simple agents that just decide that a person is in their office, in a meeting, or has a scheduled meeting, then the model of the user and the office dweller, becomes central to deciding whether or not an appropriate interruption should occur. The interruption by a person in a different part of the organization is much easier to make a decision about. This idea that simple sensors and models of the situation should be part of mediating communication in those scenarios has been very productive. In driftctcher (Lockerd,2002), an email system uses clear graphical depictions to identify and label communication by its purpose and by the value of responding to that person quickly. An experiment evaluating annotated email demonstrated that people would choose to respond more to others if they had some indication of the value of that communication.

As we look to tools that will add further to the value of such systems, we understand that the value of different communications can be automatically recognized, to some extent in order to be valuable by mediating systems. The state of these context-aware systems demonstrate that in fact even when they only partially understand what the value of the information is, we can still improve a person's ability to be productive using it. This approach to building simple, but usable Context-aware systems has the advantage that they solve well defined problems, which makes them successful at their task. The approach is based on the idea that a tool can have multiple uses. That is, tools can be adapted to solve a range of problems, and when used in this way, they are highly effective.

This approach is also based on the philosophy of using serendipitous information as a source for contextual knowledge. Our approach can be summarized as identifying the minimum sensor data that can provide us with enough information about the user to provide improvement in the user's activities. A major implication of this philosophy is that chances for adoption increase, as it is not necessary to make major modifications to existing systems, such as buying expensive hardware, sensors, or installing and configuring new technologies.

In order to show the value of the virtual sensor model being inserted between sensor and effectors, we focus on technologies that are widely spread and do not make assumptions about what type of technologies might be available 20 years from now. By identifying real problems and their current solutions, we can only hope that these same problems could be better solved with future technologies. The same problems simply would be solved with a different set of tools; the context-aware approach however continues to make the scenario more robust. Proactive and reactive agents in user interface to recognize and respond appropriately to people's needs, has for some time been a direction that has been fruitful (Selker 1994).

2.5 Implicit Metrics

Implicit metrics are a compilation of data implicitly given by the user while focusing on explicit actions required by the user-interface in order to accomplish an activity. The use of implicit metrics has been demonstrated through several context-aware examples. Attention and intention are often a driving force in these systems. Collection of implicit measures is an efficient alternative to collecting explicit user feedback, which can be costly in time and resources. Implicit behavior detection is used to find out indications of user interest. We focused on the events in computer interfaces that can be automatically captured to analyze application usage or and predict user interest.

We explore the use of mouse movement as an indicator of visual attention and a viable alternative to eye-tracking systems (chapter 7 describes implicit metrics in detail). This work focuses on mouse monitoring because it is the main interaction medium across several desktop platforms and is included as part of the interaction scenario without the addition of new sensors for monitoring user activity. The mouse can be seen as a ubiquitous universal sensor that is platform independent (Linux, solaris, windows, macs, etc.), and currently in use by millions of people everyday.

2.6 Chapter Summary

Although research using controlled environments has provided insight regarding human reactions to interruptions, it is not easily generalized due to differences in task and disruption metrics. On the practical side, sensor-based models show potential to predict user disruption; however, this approach focuses on measurable metrics and is often application dependant.

The work in this thesis differs from previous work in that we place emphasis on the user goals and motivations that influence the interruption process. Existing work has focused on local performance factors of efficiency and task success. Whereas, our approach focuses on the qualitative aspects of human computer interaction and argues that overall success relies heavily on perceived disruption. Our work also differs from existing research in that our approach takes into account the interruption content and how it relates to the ongoing task.

C H A P T E R

3

EXPLORATORY EXPERIMENTS

3.1 Introduction

This chapter discusses our initial work and findings in evaluating interruptions in human computer interaction. A series of experiments investigating the use of ambient displays as interruptions and techniques for controlling interruption are described. These studies are presented as an effort to se the context for understanding empirical work related to the human interruption process. These initial findings motivated a disruption model and experiments addressing the model.

Our initial studies explored the use of ambient displays in the context of interruption. Ambient displays acted as external interruption generators designed to get users' attention away from their current task (Arroyo and Selker 2003). The experiment investigated questions about what parameters a computer interface could use to determine the proper interruption modality to use. The experiment explored the transition in ambient displays from the background to the foreground using tactile and visual modalities. This work verified that the disruptiveness and effectiveness of interruptions also varies not just by interruption timing, but also with the modality used to interrupt.

A second exploration yielded an adaptive disruption systems designed to control and understand disruption effects from interruptions. CarCOACH is one of these explorations; it presents scheduled feedback controlled in terms of quantity of total feedback and feedback with regards to a specific stimulus, and driver current state.

3.2 Ambient Displays as Interruptions

3.2.1 Ambient Displays

In ambient displays, information is moved off the screen in a way that makes use of the entire physical environment as an interface for digital information (Wineski, et al 1998). Information is presented in subtle changes in form, movement, color, smell, temperature, or light. One example is the representation of activity by a pattern of illuminated patches projected onto a wall (Ishii, et al 1998).

Ambient displays seek to present information in the modality and form that can be interpreted with minimal cognitive effort. Ambient information is processed in the background. The person decides whether or not to move it into the foreground or back again. A person has the option to choose to focus his attention on ambient information displays at will. In the presentation of ambient media, one of the key elements is the modality chosen to present information. The choice of modality for the background media should be considered with the person's foreground task in mind.

3.2.2 Multiple Modalities

Traditional Graphical User interfaces focus only on a small number of modalities to interact with users. Finger and hand actions with the keyboard and mouse are the most commonly used channels to interact with computers. Moreover, visual and acoustic modalities are most often used for presenting information to users. These computer interfaces generally ignore important modalities such as ambient and peripheral visual cues, heat, vibration, smell and the sense of touch. Past work provides evidence that there are substantial advantages in efficiency by using multimodal interfaces (Oviat 2000), however the focus of multimodal HCI research has been on combining input modalities – such as speech, pen, touch, hand gestures, eye gaze, and head and body movements– rather than using multimodal outputs to take advantage of human sensing capabilities.

HCI could be greatly improved by using multimodal interfaces that involve the use of all human senses. The common and unique characteristics of the human senses (their ability to be ignored, precision and speed) allow for interfaces that use multiple modalities and select the modality to use based on contextual information. Even though recent work has explored the area of adaptive user interfaces, they do not consider adapting the output modality itself. The work presented in this section demonstrated that it is possible to differentiate between modalities and build multimodal interfaces by dynamically selecting the interruption modality to use, based on its effectiveness, user's performance, and disruptive effects.

3.2.3 Experimental Design

Ambient displays strive to present information in the modality and form that can be interpreted with a minimal cognitive effort (Wisneski, et al 1998). However, they can also act as external interruption generators designed to get users' attention away from their current task. Ambient displays can help orient and situate a person to serve a purpose other than the mere presentation of information—they serve as a media for creating and changing context about interruptions.

A multimodal interface was created to communicate with users through multiple channels by using several ambient displays. Interruptions were presented in the form of heat and light in order to demonstrate the benefits of using other perceptual channels in current computer interfaces. These ambient displays acted as external interruption generators designed to get users' attention away from their current task—playing a game on a desktop computer.

A lamp located at the periphery of subject's field of view (approximately at a 45-degree right to the screen) varied its intensity level (from 5% to 95%) to represent different information. A thermal copper mouse pad used Peltier devices to warm a wide area in contact with the user's hand to signal an interruption, see Figure 1. Temperature moves from ambient temperature to a warmer temperature at a rate of about 1 °C per second. The temperature range goes from 22°C to 40°C.

One of the main hypotheses evaluated is that users' performance differs based on the interruption modality. A second hypothesis stated that the perceived disruptiveness of an interruption varies depending on the interruption modality. Finally, an alternate hypothesis stated that subjects' performance is negatively affected if interrupted by their non-preferred modality.

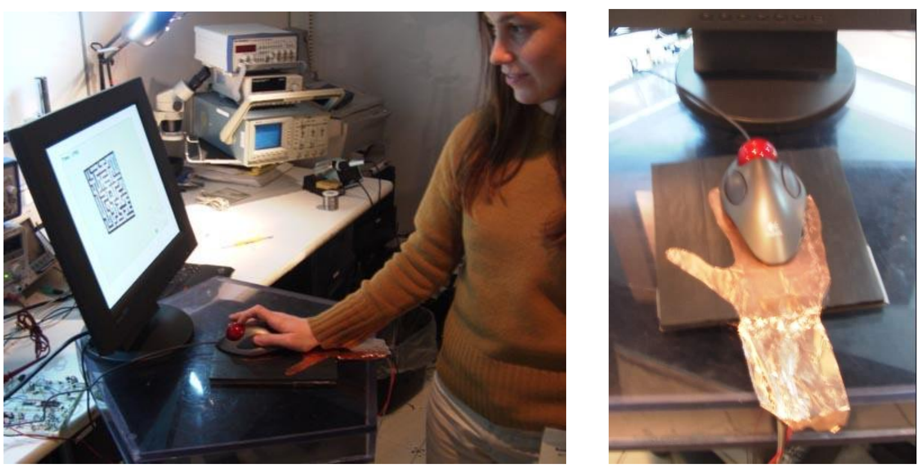


Figure 1 Tactile display prototype: A thermal copper mouse-pad warmed the user's hand to signal an interruption

3.2.3.A Experimental Task

The interruption of people during human-computer interaction is a high-level interdisciplinary topic. Interruption is a complex process that involves many subtle low-level mechanisms of human cognition (Bailey 2000-A). Therefore, a reasonably complex experimental task is used to elicit the appropriate cognitive load. It might be possible to investigate the process of interruption at the level of user interface design without fully understanding the many subtle low-level cognitive mechanisms involved

(McFarlane 2002). In this experiment, the smaller effects are ignored and isolated from the high-level effects by looking only into aspects of the human-computer interaction.

The abstract task is a simplified model of common real world tasks. The experiment is set in the context of a computer-based adventure game, similar to online Multi user Dungeon (MUD) games, where the player has to issue commands to the computer in order to achieve certain goals, Gillie, et al used a similar approach (Gillie, et al 1989). A MUD (Multi-User Dungeon) is a network-accessible, multi-participant, user-extensible virtual reality and has an entirely textual interface, see Figure 2. Participants type commands and the computer displays text corresponding to the action taken. Participants have the appearance of being situated in an artificially constructed place. An example of a kind of person performing this type of tasks is a software developer. A debugging task, for example, requires a software engineer to identify and keep track of variable values as they change over the execution of the software. A software engineer has to create a mental grid and memorize several values while looking for the next line of code to execute. These identification and tracking tasks impose a high cognitive load and interruptions during this process causes errors, allowing for observations of subjects' responses to be easily broken down into discrete units.

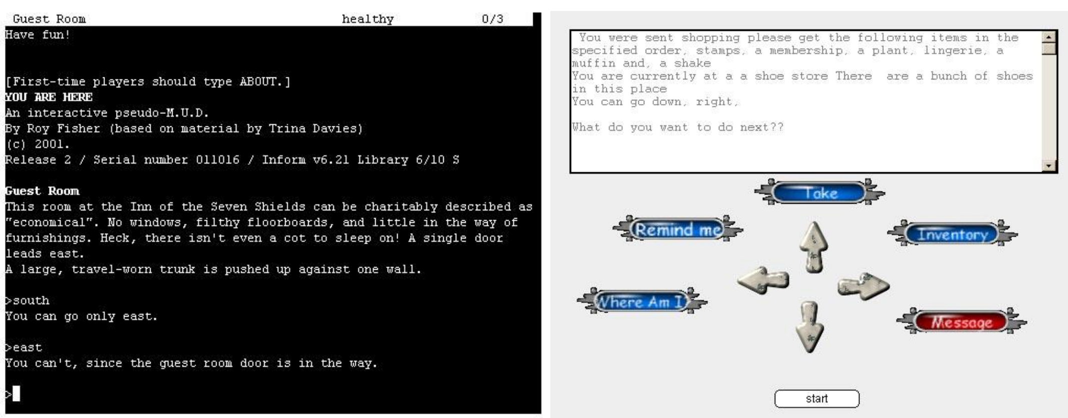


Figure 2 Text-based MUD (left) and Hybrid MUD window application developed (right).

3.2.3.B Protocol

Subjects were asked to perform the high level cognitive task involving a text-graphic hybrid of the game, see Figure 2. The task provided several performance and disruptiveness indicators: score, speed, error rate and overall time. Czerwinski also used a memory task to look for effect of interruptions while subjects navigated a list of items searching for a book title (Czerwinski 2000-A).

While subjects performed the primary task, an ambient device attracted their attention by changing temperature or by changing light intensity. Subjects acknowledged the interruption and performed a secondary task: reading a list of topically related words, similar to a free recall test. The dual-task is conceptually simple, but difficult to perform due to the high cognitive load. Several measures were collected and grouped into three main categories: disruptiveness, performance, and effectiveness. For this experiment, disruptiveness is defined as the error rate produced by the interruption modality in the primary task. Performance is defined as the time spent taking objects. Effectiveness is defined as the time taken by the user to acknowledge an interruption.

The computer game presented subjects with a series of problems; each problem containing a list of six items to be taken in a fixed order. Gillie et al compared the effect of flexible plans with arbitrarily fixed order plans and logical fixed order plans, and reported that people performed similarly across the three types of fixed plans (Gillie and Broadbent 1989). Aditionally, Miller suggests that fixed plans use more working memory than flexible plans (Miller, et al 1960).

3.2.3.C Participants

23 subjects from the MIT student body responded were recruited and compensated for their time. The sample consisted of 14 males and 9 females with ages ranging from 22 to 34 years.

3.2.3.D Results

For ease of readability, the results after evaluating the dependent variables are presented in four categories: performance, disruptiveness, effectiveness, and other measures. All of which support the hypotheses previously stated.

3.2.3.D.1 Performance

A One-way repeated measures ANOVA applied to the time to take each object after an interruption revealed that there is significant difference in performance caused by interruptions. The Huynh-Feldt epsilon was applied to the degrees of freedom to account for violation of the sphericity assumption, $F(1.6, 36) = 819.47$ $p<0.0005$. Pair-wise post hoc comparison reveals that there is a significant difference in performance for non-interrupted tasks (20.32secs. per objects) Vs. interrupted tasks with heat (32.25secs. per object) and light (25.32secs. per object), $F(1,22)= 30.89$, $p<0.0005$, $F(1, 22) = 6.47$, $p<0.19$. Figure 3 illustrates a graph showing the increase in performance for each of the modalities. As expected, performance increased due to learning effects. However, when interruptions were introduced, performance improvements diminished; particularly when using heat as the interrupting stimuli.

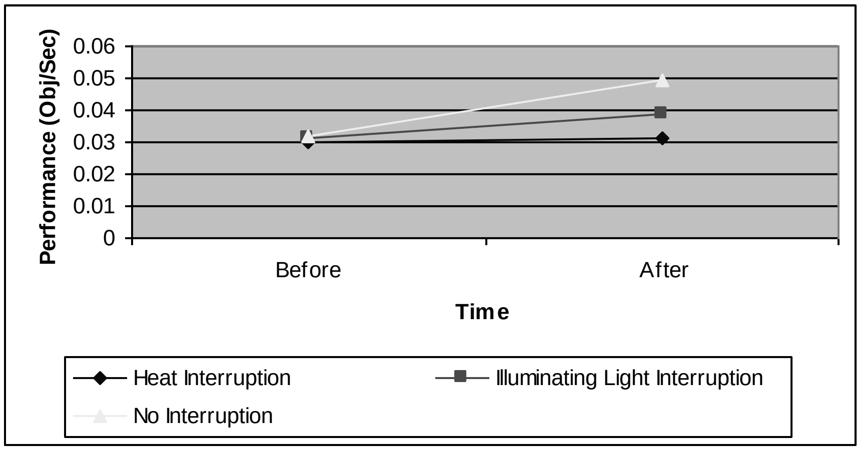


Figure 3 interrupted vs. non-interrupted. Increase in performance for non-interrupted tasks over time.

3.2.3.D.2 Disruptiveness

One-way repeated measures ANOVA applied to the number of errors in direction after an interruption for heat and light reveals that there is a main effect from interruption modality on error, $F(1, 22) = 5.478$. $p<0.029$. The error rate for heat was 0.45 errors per trial and 0.21 errors per trial for light. A one-way repeated measures ANOVA compared the number of errors before an interruption vs. after, and found there was a significant effect of interruption $F(1,45) = 19.855$ $p<0.0005$. Figure 4 shows the difference in rate of error for heat vs. light interruption.

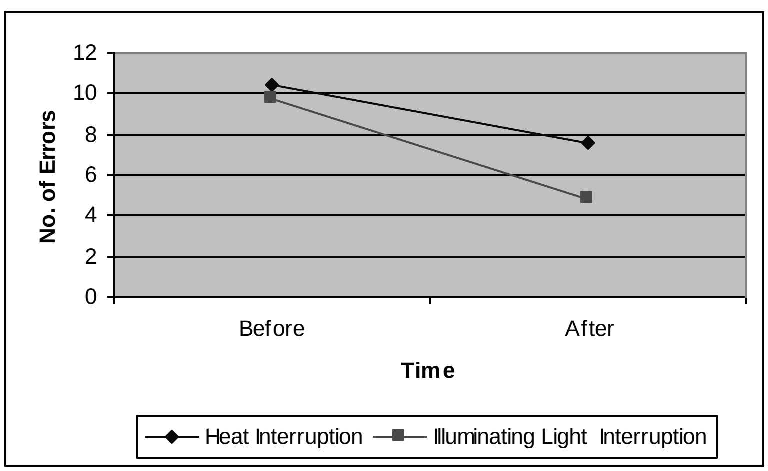


Figure 4 Difference in error rate for heat vs. light interruption

3.2.3.D.3 Effectiveness

Measures of reaction time associated with the time to acknowledge an interruption were tested for differences with one-way repeated ANOVA. The analysis indicates that there was a significant difference in reaction timed for heat (9.60secs.) and light (5.50secs.), $F(1, 22) = 7.76$, $p<.011$.

3.2.3.D.4 Preferred modality

Subjects were asked to choose their preferred modality subjectively. 40% of the subjects selected heat as their preferred modality, and the remaining 60% of the subjects selected light. One-way repeated measures ANOVA shows there was no main effect of preferred modality in performance difference, $F(1,22) = 1.374$, $p>.254$, and neither in speed, $F(1,22)= .006$, $p>.94$.

3.2.3.D.5 Observations

One surprising comment about heat was the fear of being hurt. In general heat was perceived as a dangerous threat. It was also generally mentioned that heat is slower than light, and thus harder to detect. Interestingly, although heat was harder to detect, it was also harder to ignore once it was present. This is probably because subjects associate heat with danger, and consequently, did not dare ignore it, anticipating being burned. Alternatively, light could be postponed until the active task had finished. Light, as opposed to heat, which had an affective component, had no physical interaction with subjects that could be perceived as an invasion to their personal space. 39% of subjects agreed that light interruptions are easier to identify than heat interruptions. There were mixed comments about how disruptive light is, some mentioned light is more disruptive and others mentioned light is less disruptive. There were only 8% of subjects classifying light as pleasant. There were mixed comments about how disruptive heat is, 50% of subjects classified heat as more distracting or disruptive, whereas the other 50% classified heat as less distracting and less obtrusive. Some subjects even mentioned heat as pleasant, especially in cold environment or as an aid for carpal tunnel syndrome treatment.

3.2.4 Conclusion

This experiment verifies and extends previous research about interruptions, in that subjects perform slower on an interrupted task than on a non-interrupted task, demonstrating the general effect of interruptions. Furthermore, this experiment also shows that the interruption modality affects performance. The thermal display produced a larger decrease in performance than the visual display. This thermal display also has a greater disruptive effect on the interrupted task than light. Disruptiveness and performance measures agree that heat causes more of a detrimental effect than light when used as an interruption.

The results provide information about how to use heat and light as modalities for interruption. The work showed that light is more efficient (42% faster than heat) in getting users' attention, but has a disruptive side effect on speed slightly larger than heat. In contrast, heat takes longer to be noticed. Heat could be used more reliably in environments where other channels are already saturated or overwhelmed with information, such as when there are many visual distractions. One of the advantages of using heat is that users can attend to an interruption without having to take their attention off the screen, whereas with light interruptions, users tend to focus their attention to the light source. Heat acts as an interruption to a single person, without disrupting everybody around them. Ambient lights, however, alert all people present at the location where light changes occur. Heat can be used to signal messages subtly to a single person, that is, heat is a personalized attention-getting device. Light can be used to signal messages when there is only one person present at certain location, or when multiple recipients are intended.

These results set the initial point in motivating the theory behind future self-adaptive interruption interfaces that will employ users' individual responses to interruption and dynamically select the modality based on effectiveness and performance metrics. This work suggests that computer interfaces can arbitrate interruptions based on their disruptive effect. Thus it is conceivable to maximize the effectiveness of an interruption through proper notification configuration, selection, and timing.

3.3 Mediating Disruption in real-world settings

This section presents data demonstrating the value of a context aware driving advisor. It describes the design, implementation and evaluation on the road of a car-integrated system designed to promote better driving behavior. CarCOACH is a system built to explore and demonstrate that the COACH approach could help tech people in a dynamic environment of distractions. The system was developed in collaboration by Shawn Sullivan, Taly Sharon and members of the context-aware computing group. In the following exploration, vibration and audio interruptions were shown to be plausibly non disruptive when context played a central role in scheduling interruption.

The Cognitive Adaptive Computer Help (COACH) approach advises people to enhance their understanding and skills (Selker 1998). This approach demonstrated that choosing the right amount and style of information to display while a person is programming could greatly improve their ability to learn and use programming constructs. This style and approach of choosing how much and which communication is important was motivational for the driving advisor system presented in this section.

The high levels of computerization in cars allow us to monitor several aspects of vehicle activity. Activity data already represents much of the driving behavior if interpreted that way. One of such examples are black boxes that allow parents to track where the car has been and its driving conditions. Rather than just being, a logging tool CarCOACH uses sensors together with models of driving behavior to monitor driving successes and mistakes. It uses these models to predict possible times to intrude and present feedback, reminding people to drive their best. CarCOACH works on real environments, presenting direct feedback at the point of behavior (driving a car) by employing tactile and auditory interaction modalities.

3.3.1 CarCOACH

CarCOACH is a car-integrated persuasive system that presents "just-in-time" (Fogg 2002) context-sensitive feedback to users with the goal of reminding users of appropriate driving techniques and promote behavior changes regarding driving habits. CarCOACH interprets standard driving controls in the car (RPM, speed, throttle position, brake pressure/position, steering position angle, cup holder state, and on-board system status). CarCOACH is able to identify common driving behaviors, such as excessive braking force, extreme acceleration, turning without signaling, driving erratically, and excessive turning speed. CarCOACH predicts constructive times to intrude and present direct feedback.

Support and warning systems have been developed to assist un-educated drivers (Michon 1993). These systems suggest that technological solutions can provide feedback on driving ability, warn about dangers, and ultimately improve driving performance (Hutton, et al 2001). CarCOACH is an artificial intelligence application that uses the blackboard architecture model (Jagannathan, et al 1989) to implement conflict resolution to provide feedback while driving. The system presents scheduled feedback controlled in terms of quantity of total feedback and feedback concerning a specific stimulus, and driver current state. The schedule's goal is to reduce driver's stress and maximizing the effectiveness of the feedback presented.

3.3.1.A Design

CarCOACH is a persuasive system that presents "just-in-time" context-sensitive feedback to users with the goal of inspiring behavior changes regarding driving habits. CarCOACH's design is based on behavioral modification theories and seeks to encourage users to gradually change their habits. Application of persuasive theories to user interface design for ubiquitous computing has been shown to be very effective at generating sustainable changes in behavior (Cialdini 2001; Fogg 2002, Skinner 1991).

The point of behavior for driving cars is the car itself: People driving on the road and performing common driving tasks. Since the point of behavior is the car, CarCOACH was developed on the 300M IT-Edition, a concept car built through collaboration with MIT Media lab and Chrysler. It is an instrumented research vehicle equipped with many additional driver monitoring sensors, data loggers, and network capabilities (Pompei, et al 2002). The use of a real car accounts for the fact that simulators do not provide a vivid experience as compared to real driving conditions (Gibson 1998). CarCOACH is also designed to improve the techniques of skilled drivers by giving appropriate proactive feedback, such as vibrating the throttle pedal when users accelerate abruptly. CarCOACH is an agent designed to give the driver appropriate feedback on his driving performance. CarCOACH design follows the following principles for in-vehicle information systems:

- Does not distract the driver
- Does not result in unsafe driving conditions during total or partial system failure
- Does not present a hazard as a result of unintended or naïve use by inexperienced users
- Its displays do not aim to visually entertain the driver
- Does not produce patters or sounds liable to startle the driver
- Does not require training—it is easy to use
- Is able to be switched off without affecting the control of the vehicle

CarCOACH presents direct feedback using non-obtrusive interaction modalities in the form of tactile and auditory reminders. These modalities allow the system to present information without interfering with the user's task at hand. The auditory reminders refers to messages presented by a female voice saying "thank you for signaling;, "good breaking", "good acceleration", "good braking", "easy on the gas", "please drive smoothly", and "please signal". Tactile reminders include steering wheel vibration and break and acceleration pedals vibration.

3.3.1.B Hardware Architecture

An embedded Ethernet controller solution provides sensor data streamed out using UDP protocol over an Ethernet network available throughout the car. Individual boards perform signal and distributed data processing. A Vehicle communications network interface (SAE J1850) monitors RPM, speed, throttle position, and on-board system status. Additional custom-fitted sensors include brake position, cup holder sensor, and steering position angle.

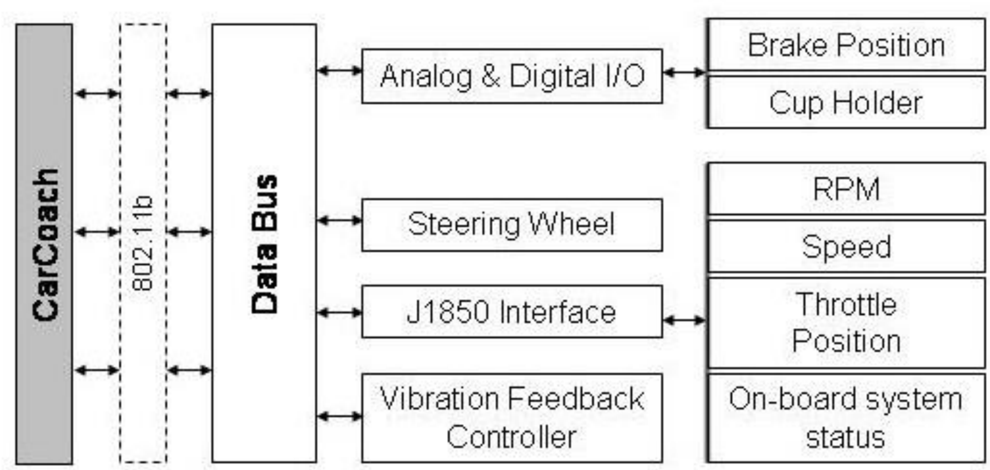


Figure 5 CarCOACH Hardware architecture. The J1850 standard , is a low cost, master-less, and single-level bus open architecture.

The system architecture supports wireless 802.11b to separate computer used to collect and interpret received sensor data, see Figure 5. Each sensor input is continuously monitored, and its signal entropy is calculated. When these values cross certain thresholds, the system determines what the stimulus was and updates knowledge sources on a blackboard coordinator.

3.3.1.C Software Architecture

The purpose of the data analysis and feedback system of CarCOACH is to take the information gathered from various sensors in the car, interpret it, and make decisions about informing the driver of mistakes and/or correct driving, ensuring that drivers do not experience sensory overload and are not given feedback in a dangerous situation.

The data analysis and feedback system design is based on the Blackboard Architecture. The blackboard model allows multiple independent agents (knowledge sources) to share information in a central scheduler store (Jagannathan 1989). At the center of the system is a mediator which analyses data from several knowledge sources and then decides which knowledge source should be given priority at the time, Figure 6 illustrates the mediator, scheduler and feedback interactions.

Agents in the system monitor data related to a single aspect of the system. They monitor acceleration, braking, turn signal use, turn speed, and how erratic the driver's steering is. Further, the data is normalized in such a way that a greater magnitude indicates absolutely a greater current significance to the system than any other agent whose data normalizes to a smaller value. For example, if the acceleration agent produces the largest normalized data of all agents, then the acceleration agent might be the most significant agent in the system at that moment. Data normalized this way is called an attention score, since it directly represents the amount of attention an agent, and hence an aspect of driving, currently requires. The agents also have the ability to respond the driver's actions; that is, all feedback is generated by a single specific knowledge source. In addition, there is a scheduler for providing feedback.

The purpose of the scheduler is to prevent sensory overload. Feedback should not always be given to the driver for every stimulus, as this would quickly become a cumbersome and overbearing system for the driver. A scheduler decides whether feedback is appropriate in terms of quantity of total feedback and feedback concerning a specific stimulus. The scheduler can also delay feedback in situations in which it may be too dangerous to provide immediate feedback (Sharon 2005). When feedback is deemed appropriate by the scheduler, the corresponding knowledge source is asked to respond. To ensure feedback is not given in a dangerous situation, the scheduler examines the attention scores of the agents, and if they are so high that the driver might be in a dangerous situation, such as very high braking and turning speed scores, potentially indicating a skid, the scheduler may alter the timing of feedback. If the scheduler determines that immediate feedback may cause sensory overload or that the current situation is too dangerous, it may delay or cancel the feedback.

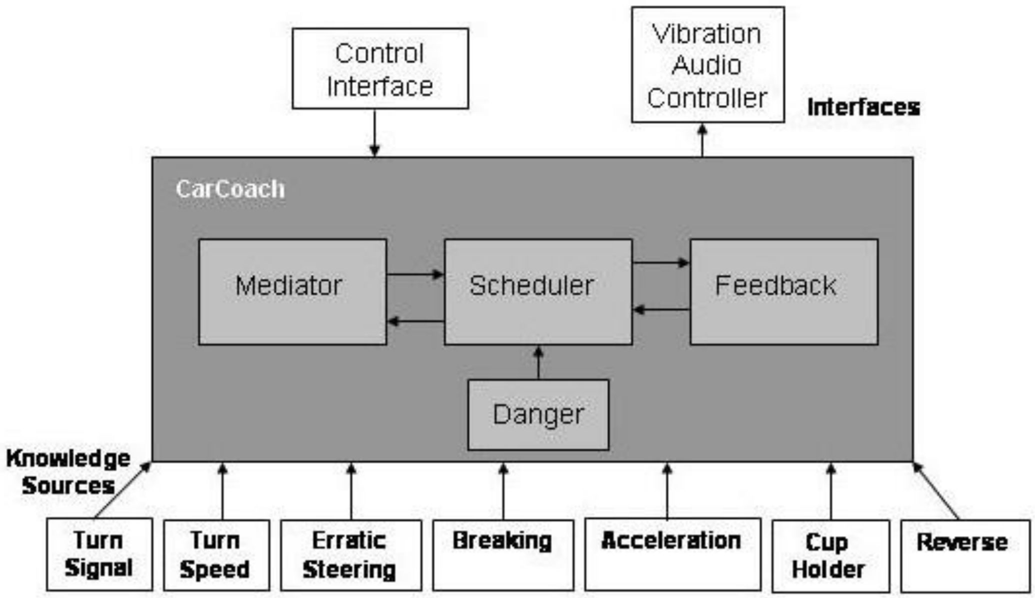


Figure 6 CarCOACh Blackboard Software Architecture includes seven knowledge sources.

A prototypical knowledge source contains information universal to all knowledge sources, including the score and the number of types a positive or negative feedback has been given from that knowledge source. Five knowledge sources monitor acceleration, braking, erratic steering, turn signal and turn speed. Two special knowledge sources have the ability to affect the busy state of the car, rather than providing feedback. These prototypical knowledge sources monitor whether a cup is in the cup holder or may be in the driver's hand and whether the car is in reverse as safety precautions.

The Scheduler looks not only at the quantity of feedback given to the driver but also a danger score which will not allow any feedback while the driver is still in danger. When a knowledge source indicates it should provide a response to the user, it does not actually respond until the Scheduler clears the attempted feedback. Feedback should not be given in situations in which the driver is in danger; if feedback is required in a dangerous situation, it should be delayed or canceled.

3.3.1.D Feedback Generation and Scheduling

A description of the architecture serves to enlighten the technical details of CarCOACH, and some sample hypothetical scenarios help to show how it behaves in real-world situations.

When driving is within normal parameters, the system provides no feedback. However, suppose that the driver suddenly presses the brake pedal too hard. The agent monitoring the break pedal increases its attention score in concert with the increase in brake pressure to a level above the threshold for needing attention, telling the central mediator that the brake has been pressed hard. Assuming no other agents increase their attention scores, the mediator then decides that the brake being pressed too hard is the major problem, and interacts with the scheduler to arrange for feedback to be given. If the driver has not been given too much feedback about braking or overall lately, and the situation is not dangerous, then the driver receives feedback.

However, suppose the driver has been making a lot of mistakes lately and been receiving a lot of feedback. To reduce the likelihood of cognitive overhaul and frustration with the system, the scheduler will postpone or cancel feedback. Feedback is rarely postponed due to this reason, as by the time the fears of cognitive overload have past, the driving situation is changed. Only when two separate events requiring feedback occur within seconds of each other, such as rapid acceleration followed by braking, is the feedback postponed, since the driver will hear the audio feedback nearly sequentially and easily be able to connect his actions with the feedback received. Otherwise, as is usually the case, the feedback is canceled to prevent confusion with what exact the feedback concerns.

In another situation, the mediator may decide feedback is warranted, but consultation with the scheduler shows that immediate feedback may be dangerous. For example, the driver may have pressed the brakes so hard he is in a skid and therefore immediate feedback may only prove distracting. To assess the danger level of situations, the scheduler examines the attention level of the agents, and also factors in other characteristics such as driving speed steering wheel position. If speed is very high and the agent monitoring the brake pressure has a very high attention score, then the situation is very dangerous, more dangerous than if the speed were very low and the brake pressure very high. Likewise, a sharp angle on the steering wheel combined with a high attention score from the agent monitoring the acceleration of the car may also be dangerous. If these or other similar dangerous scenarios are the case, then the scheduler will not allow any feedback to be given to the driver until the system is no longer dangerous.

The scheduler also ensures that no feedback is given to the driver while he or she might be unusually busy with a particular task not generally performed while driving. For example, a sensor indicates if the car is in reverse, then the scheduler delays or cancels all feedback, because the driver is looking behind the car and probably concentrating more than usual.

A final situation CarCOACH may be involved in is a dramatic event such as an accident. As the accident is happening, CarCOACH will continue to provide feedback in an effort to prevent the accident if the scheduler does not believe the situation is too dangerous. However, if the situation is too dangerous and the accident cannot be avoided, no feedback will be given because the driver has enough to concentrate on in trying to avoid or prepare for the accident.

3.3.1.E Interface Considerations

CarCOACH is based on an understanding of existing and emerging theories and strategies for motivating people to behave differently, with the goal of getting users to drive their best more often. Studies have shown that easy-to-understand messages displayed at the right time and place in a non-annoying way can influence behavior change (Intille 2003). Prompts are an effective technique to encourage sustainable behavior because they remind people of actions that they are predisposed to do (Russell 1999). Computer technologies now make it possible to deliver reminders and prompts right at the point of behavior in response to user activities.

Researchers from a wide range of fields have demonstrated the value of using "just-in-time" prompts to engage sustainable behavior change (Aronson 1982). CarCOACH presents "just-in-time" context-sensitive feedback to remind users of appropriate driving techniques at the point of behavior; when a mistake is made. CarCOACH's feedback takes into account the amount of previously presented feedback in order to reduce the stress and frustration levels from using the system. Auditory feedback in the form of descriptive messages acts as prompts, reminding users about their current action and appropriate techniques to use while driving. Auditory messages were chosen because they are easy to interpret, are noticeable, and do not interfere with the driving task at hand.

A positive reinforcer is anything that a user desires and that occurs in conjunction with an activity. If presented at the time of an action, a positive reinforcer tends to increase the likelihood that the action will be repeated (Cialdini 2001). The system uses positive reinforcement in the form of audible messages in conjunction with vibration directed at the device involved (steering wheel, brake pedal, etc.). An appealing message plays when the user performs a correct driving maneuver. While constant reinforcers must be presented at the beginning stages of behavioral modification, once the desired behavior is established, they become less effective and could potentially become annoying. In order to maintain newly established behaviors, CarCOACH has the ability to schedule the reinforcement, reinforcing a good driving maneuver occasionally and on a random and unpredictable basis.

3.3.2 Experimental Design

Our primary goal was to test whether CarCOACH had beneficial effects on driver's performance. An exploratory experiment examined how effective CarCOACH is at improving driver performance when compared to continuous feedback and no feedback at all. The experiment also examined (user acceptance) how users would react to a systems presents feedback politely while providing feedback on the road.

We hypothesized that driver's performance would be reduced when constant feedback is presented and that CarCOACH would elicit higher driving performance when compared to continuous feedback and no feedback. We also hypothesized that CarCOACH would be perceived as less disruptive and polite than a system that presents continuous feedback

3.3.2.A Protocol

Subjects were randomly assigned to two conditions. One condition presented continuous positive feedback at the beginning of the trial and scheduled feedback at the end of the trial. The other condition presented scheduled feedback first and ended with continuous feedback. Half of the subjects were assigned to each condition. Before the experiment, subjects performed a test drive for 20-30 minutes to acquaint users with the driving characteristics of the car. During this test drive, the system collected baseline performance metrics in order to calibrate the feedback presented by the system at later stages. Subjects were presented with feedback in the form of sound and vibration for half of the trial. For the second part of the trial (after a five-minute break), subjects continued to drive for another twenty minutes and were presented with a different feedback condition.

A one-way repeated measures design with three feedback conditions as the independent variable: Baseline, Continuous and Scheduled. User's performance was used as the dependent variable—CarCOACH continuously monitors driver's performance and assigns a Score based their performance. Additionally, CarCOACH internally assigns rewards when good driving is identified (driving on target). This value was also used as another metric for performance.

3.3.2.B Results

Planned comparisons for performance measured by score indicate that performance for continuous feedback decreased 49.4% from the baseline performance $F(1,8) = 6.99$ $p < .05$. Not being considerate about when to give feedback actually reduced performance. Whereas performance using scheduled feedback through CarCOACH increased 34.7% than the baseline score, but had such variance with our short trials and small subject size that it was not significant $F(1,8) = 2.35$ $p = .16$; $r = .5$.

The effects of positive feedback are better observed over longer period of times. The Performance under scheduled positive feedback was significantly better (166.4%) than performance under continuous feedback $F(1,8) = 33.57$ $p < .001$. One-way repated ANOVA indicates there is a main effect for performance as measured by driving on target score $F(2,16)=46.62$ $p<.001$. Post-hoc comparisons indicate scheduled feedback yielded a significant 56% increase on performance $F(1,8)= 42.3$ $p<.001$.

This exploratory experiment shows that CarCOACH has the potential to increase performance. Comparisons of performance levels suggest that CarCOACH performs better than a system with continuous feedback.

3.3.3 Conclusion

With this work, we were able to test whether an interrupting system had beneficial effects on performance. In addition, a variable schedule of feedback for remarks about driving behavior was compared against continuous feedback and no feedback at all. A controlled experiment examined the effects of feedback type and scheduling schemes on driving performance and elicited frustration. This driving experience showed considerable differences between positive, and negative style feedback. Negative feedback proved to be the worst by decreasing performance significantly eliciting high frustration levels on drivers. This negative effect can be reduced by scheduling the feedback presented and introducing scheduled positive feedback. The results indicate that if negative feedback needs to be presented, scheduling feedback could reduce its higher frustration.

This project showed insight that even in a car; drivers' performance can be improved by giving advice now and then to make them aware of their driving performance. The variable schedule of feedback was more affective at increasing driver performance. Adaptive agents can help people by replacing their actions with assistant or teaching and improving their performance with advisory agents.

This worked presented evidence that advisory systems can be used in complex environments competently The use of agents that use models of task, system and users can be used together to make a system that competently knows how to give encouraging feedback. Feedback can be considered an interruption and, as such, it can be disruptive to the driving task and be perceived as disruptive. However, the balance between disruption and performance benefits cannot be easily identified. In order to make a theory that encompasses this important ideal we must extend current interruption-disruption models.

3.4 Engagement Tracker

Another system we built to explore interruptions in the context of peripheral displays is "Engagement Meter". By tracking information updates, we demonstrated a scenario where a user interested in browsing the web looking for work-related information might also be interested in keeping track of a live baseball game. A disruption mediator tracks both activities and decides when the baseball video source is allowed to interrupt the primary goal. The disruption mediator observes mouse movements as the user navigates through several web pages, detects reading activity and identifies decision times, such as the user hesitating between several links. On the other hand, the mediator scans audio participation levels on a video stream, identifying potential interesting plays during the game.

The mediator signals interesting game plays by controlling the video window size in order to minimize disruption effects and allow the user to retain control over their attention focus. Increasing window size is less distracting than moving or toggling windows (Bailey 2000-B). Interruption timing depends on a combination of an interesting play being found and user activity on the current webpage.

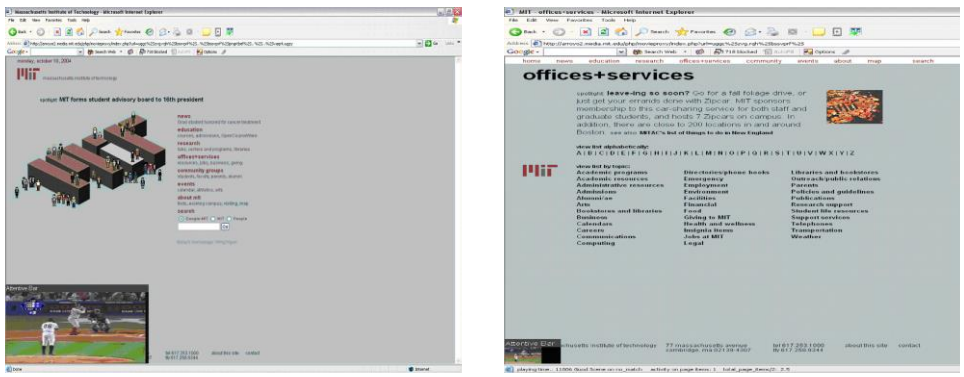


Figure 7 Video window size changes due to increase activity on webpage and web page structure. Interesting plays presented on a large window size (left) and minimum window size for intensive website activity (right)

In this baseball scenario, the user goals were clearly defined and known before hand, thus the system was capable of arbitrating between two known goals: keeping track of event highlights and locating information. Finding information was seen as the primary task and the baseball game as the interrupting task. To take appropriate disruption further, we can expand the mediator to include unknown goals by tracking web browsing activity and finding information about the topics being searched for. Since internet browsing main usage is often information search and retrieval, we can assume that the user is interested in finding information. This information supports several ulterior goals; news for leisure, games for entertainment, scholarly articles for work, and so on. Therefore, the information itself is directly related to user goals. By determining the type of information the user is looking for, we can compare the relationship between incoming interruptions and the user goals.

3.5 Chapter Summary

Recently, some progress has been made toward understanding the systems that support multitasking. Notification systems for example, (IM systems, status updates, email alerts and news, etc) attempt to deliver current, important information to the computer screen in an efficient and effective manner. Research has focused on investigating the costs, benefits and the optimal method of presentation for notifications. However systems that adaptively adjust their behavior regarding interruptions are scarce. Work has focused on understanding the effects of interruptions on dual-task situations and the relationship of the interrupting task to the ongoing task. While these efforts are useful for developing HCI guidelines for task-specific applications, an approach that acts as an intermediary between these guidelines and the needs from applications is needed.

The work presented in this chapter illustrates how we have explored the disruption problem from different approaches. First, we designed an experiment to test the effect of different interruption modality on task performance. We evaluated non-traditional presentation modalities, such as heat, vibration, light changes, and smell and found that there was a lot of individual differences in the way people reacted to interruptions depending on their familiarity with each of the modalities (Arroyo and Stouffs 2002)

On a follow-up experiment described in this chapter, heat and light were identified as less disruptive (or more subtle). On this experiment, we identified situations in which heat and light modalities could be used. These experiments motivate two research directions. One is research with the goal of providing guidelines for notifications systems that use several interrupting modalities. The second direction is the idea of dynamically controlling the modality based on people's reactions to notifications. .

A different approach to the disruption problem includes the research demonstrations presented in this chapter showing several examples of modulating interruption to reduce disruption. CarCOACH illustrated how low-level sensors had to be summarized into higher-levels activities, such as driving, steering, dangerous situations, etc. Even for driving, it was clear that the goal played a significant role. Engagement tracker also illustrates that by understanding the user goals, it is possible to balance information in order to reach those goals. Thus, we decided to investigate the role of user goals in the disruption management process. Furthermore, in order to create a more formal theory on disruption management, it is necessary to directly experiment with the interruption-disruption phenomenon itself. This thesis ties ongoing behaviors, task actions and goals with the disruptive effects of interruptions on ongoing computing activities. The end goal is to aid the design of dynamic systems that deliver efficient and effective interruptions without causing unwanted perceived disruption.

CHAPTER

4

INTERRUPTOIN-DISRUPTION MODEL

4.1 Introduction

The previous chapter presented preliminary explorations in evaluating and designing systems considerate to the interruption process. While each study provided some insight about the interruptions process, they also reveal the need to focus on a general theory of disruption as a result from computer interruptions.

This chapter introduces an abstract interruption model of human information processing that is motivated by our experiments and other research done in the field. This model is presented to serve as a guide to think about the interruption process. The chapter presents general exploration of disruptions through the model and illustrates some implications for human computer interaction design. Finally, the chapter motivates the need to evaluate the effects from interruption relevancy, task level, and task commitment to the users' goal and task priority on perceived disruption.

4.2 Task Flow in Cognitive Architectures

This section presents representations of task flow that are tied to cognitive architectures. These representations are based on a survey of theories of human information processing and cognitive architectures (Barnard 1999-2000).

Abstract interruption representations allow for exploration of the interruption process, identification of the areas prone to disruption and recognition of the parameters that might affect the interruption-disruption process. A cognitive model representation includes abstractions of mental processes involved in reacting to interruptions. Abstractions include perceptual processors, executive processes and their associated working memory and attention resources, long term memory stores, and planning. Modes of interruption are identified and specified with properties on the several factors in the model in order to predict user disruption. The model treats interruptions as part of a high cognitive level of information processing with limited mental resources, and with a control mechanism that oversees the encoding, transformation, processing, storage, retrieval and utilization of information. (Atkinson and Shiffrin,1968; Nachshon 2000; Rubinstein 2001, Altmann 1998).

On this model, interruptions are seen as a request to switch to a new or previously executed task (McFarlane 1999). The way in which this request is processed is extremely related to multitasking and task switching. This model assumes that people can execute several cognitive processes at once (multitasking), but that they can only perform one activity at a time with full conscious control and awareness. Therefore, multitasking must execute the remaining tasks as subsidiary or peripheral processes. Each of these processes has mental resources assigned to it. Thus, the mere fact of multitasking requires mental resources to evaluate and prioritize the tasks.

Mental resources assigned to an ongoing task include memory and attention. Working memory contains information relevant to the ongoing task and its goals; information retrieved form long term memory with declarative knowledge. According to the goal-activation model (Anderson 1998) information in the current attention focus is the chunk that is most active in working memory. Thus, a goal that is retrieved from memory with greater frequency will not suffer activation decay and will be easily accessible when recovering from an interruption (resumption of the primary task). The information processing model presented by Wickens (see figure Figure 8) includes several processing stages that that allow mapping of various interaction trajectories of interruptions, aiding in the understanding of the interruption-disruption process.

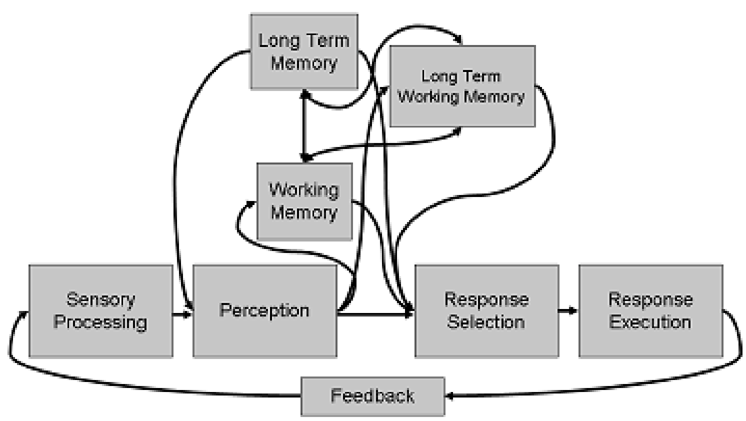


Figure 8 Wicken's human Information processing model (Wickens 2000). People use different flow of attention depending on the type of interruption.

Processing interruptions requires flow of attention to the long term memory in order to link new information to existing knowledge in order to select the desired response to be executed. Mapping the interaction trajectories of interruptions indicate that different cognitive processes are involved in attention reallocation. This mapping also shows that arrows in the information processing model depicting attention flow, will vary depending on the nature and disruptiveness of interruptions. Interruptions generate disruption and resetting of working memory, which can be seen as a side effect of context switching—such as attending to unfamiliar or complex information.

4.3 Condensed Interruption Model

Several interruptions models have been proposed. We have summarized these models into a condensed model that provides a structure for organizing basic research theory and empirical results for the purpose of better understanding the nature and effects of interruptions.

A condensed model provides a description of the processing stages involved while dealing with interruptions. In its most abstract form, the model is formed by detection and interpretation stages, an evaluation and planning stage, and an execution stage (see Figure 9). The model describes the process of detecting an interruption by means of sensory stimuli. Once detected, the interrupting stimulus is mapped to interpretations in memory and mental resources are assigned to it. A decision process will evaluate the interruption content with respect to the ongoing (system) state, plan and act upon the interruption.

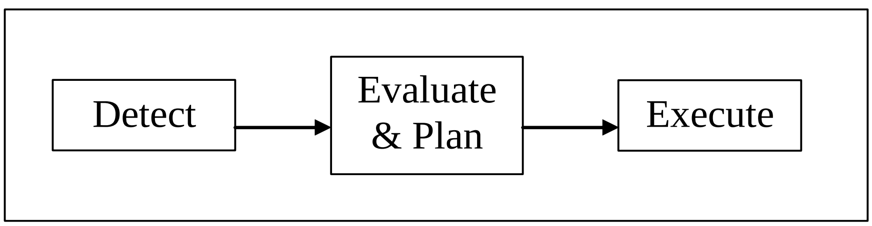


Figure 9 Simplified interruption processing model stages: Detect annunciation stimulus, interpret stimulus in terms of the interrupting task requirements, and integrate the interrupting task with the ongoing activity

Previous models, such as the one provided by Rasmussen show the cognitive processes involved in disturbance management for control monitoring and controlling processes (Latorella 1998; Rasmussen 1968). On the Rasmussen model (see Figure 10), operators observe the status of the system and use their mental model/representation of the process/task to identify the current state of the process and the cause of the disturbance. Operators later evaluate the possible consequences with reference to operational goals and define a target system state. In addition, a high level task and detailed procedure to transfer the system to the target state is defined and executed. This model identifies automated responses as being intimately related to detection of the interrupting stimulus. The model also differentiates the response to frequently observed parameters, as is the result of practiced behaviors.

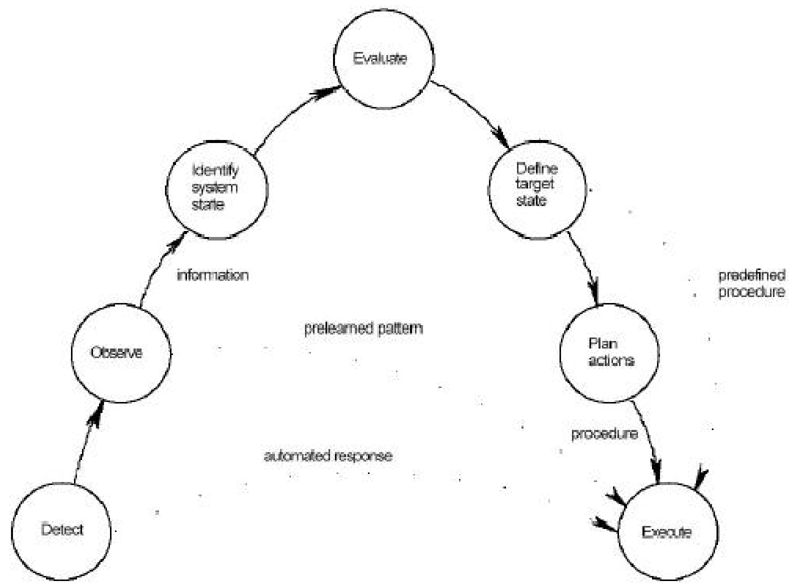


Figure 10 Cognitive processes of disturbance management model (Rasmussen 1968).

4.3.1 Interruption Detection Stage

Interruptions have an alerting stimulus characterized by changes in the environment or interface. The alerting stimulus can be visual or auditory cues, alarms, tickers, or combination of stimulus. Changes in the environment are perceived by people's external situation awareness and are further interpreted (see Figure 11). External situation awareness refers to a continuous extraction of environmental information that is fed into a perceptual processor (Adams 1990). This processor is responsible for filtering stimuli and categorizing cues extracted from the environment trough different sources (visual, auditory, etc). Situation awareness is a concept based on the notion that the people use available data sources to create and maintain a mental representation of the ongoing situation. This mental representation, the goals, objectives, and people's expectations or preconceptions determine their actions. Situation awareness decreases when task interruptions occur (Altmann 2000). Altman demonstrated that situation awareness can degrade if users don't have enough time to pay attention to the ongoing task or enough time to let previous tasks fade from memory. This suggests that interruptions should redirect attention and awareness to process events/tasks that are significant to the user's goals.

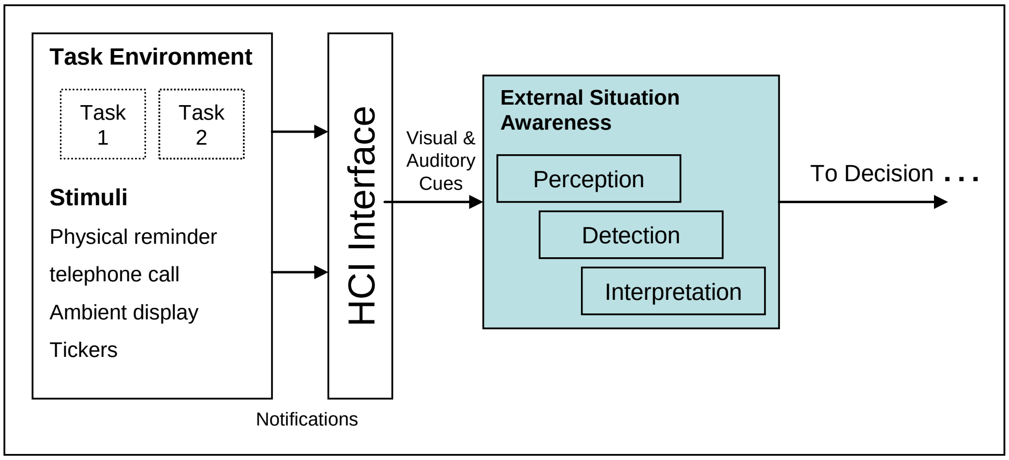


Figure 11 Changes in the environment are perceived by people's external situation awareness and are further interpreted

4.3.2 Interruption Evaluation and Decision Stage

Interruptions are accompanied by information related to the interruption and a task that should be performed. An instant message from a colleague can ask the user to check the update on a joint project on the company's website. The interruption is related to a joint project and requires an action. Interruptions can also be of the type of a notification with the sole purpose to deliver current information to the user. Such as, stock tickers or news. Notifications do not carry an associated interrupting task, but might elicit the user to engage in a new task depending on the relevance of the message.

The process of evaluating and managing interruptions include mental executive processes that manage multiple tasks and determine the priorities of each interruption to decide when to execute the task associated with it. Incoming interruptions are evaluated with respect to ongoing processes goals and priorities. Executive processes are responsible for controlling and coordinating the execution of goals, the allocation of attention to specific ongoing processes, manipulating the contents of a set of storage and rehearsal buffers for ongoing tasks (Nachshon 2000; Rubinstein 2001). Executive processes are also involved when the interference between two competing alternatives must be resolved, inhibiting attention to the irrelevant one.

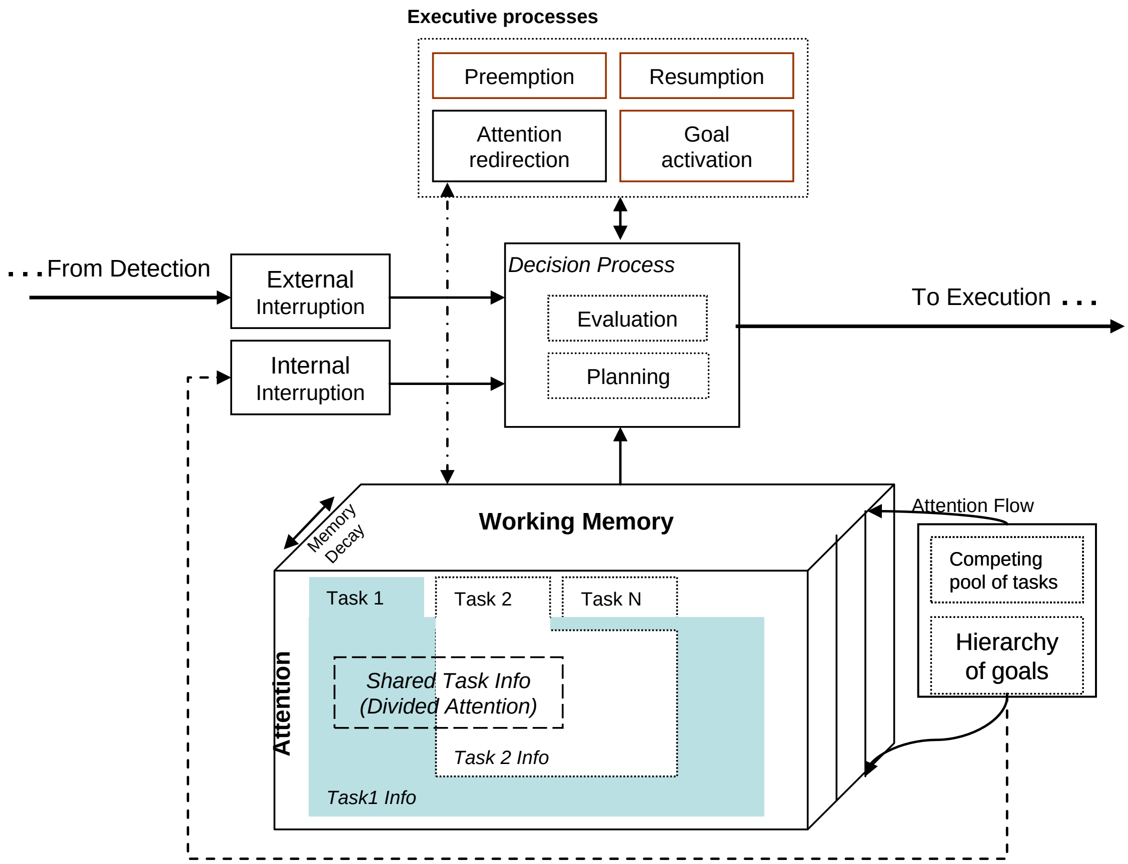


Figure 12 Decision stage in the interruption model shows abstractions of mental processes involved in reacting to interruptions

4.3.2.A Planning

The decision process must weight the benefits and cost of performing the interruption and decide whether to process the interruption immediately or to schedule it for future processing. It must preempt the ongoing task and prepare a plan for future execution considering the availability of resources. Planning involves adding a new task to the list of ongoing. This new task is responsible for triggering an internal interruption when is time to process the original interruption (Garcia-Ogueta 1993). Thus, adding to the mental load. The planning component of the decision process has a complementary set of processes available. Preemption, goal activation and resumption are useful in coping with incoming interruptions, reducing disruption, and task switching. Evaluation and planning (decision) processes are also influenced and aided by several subcomponents that provide the necessary information to properly evaluate interruptions and decide future assignment of resources: pool of tasks, hierarchy of goals, long term memory stores.

4.3.2.B Pool of tasks

The incoming interruption is evaluated against a competing pool of ongoing and projected tasks, which are contained in a set of storage and rehearsal buffers. When a decision is made to switch to the interrupting task, the ongoing context task is updated, saved and added as an additional task to the pool of tasks. Internal interruptions are originated within the pool of tasks. An internal interruption is a request by a subsidiary task in the pool of task to be switched into the person's focus of consciousness; such as switching back to the original interrupted task. (Garcia-Ogueta 1993, Kieras 1997, McFarlane 2002).

4.3.2.C Structure of goals

A network of goals classifies ongoing tasks and helps determine the priority assigned to the interruption. This hierarchy changes over time to reflect task completion and goal changes based on new information available (Baddeley 1996)

4.3.2.D Long term memory stores

Experience also influences the decision process. If one or more of the ongoing tasks have become automatic, execution might be performed immediately. Whereas if people are unfamiliar with the interrupting task, it would take longer to be processed and more disruptive (Allport 1980, Meyer 1997). Declarative knowledge related to the interrupted task is often retrieved from long term memory, such as past experiences available on long term memory stores. Research shows that experience handling interrupting tasks reduces the disruptive effects of interruptions over time (Hess and Detweiler 1994).

4.3.2.E Preemption

Preemption activates a set of mental structures in anticipation of needing them at a later stage. The mental structures include a working memory representation of the interrupted goal or task. Preemption is equivalent to marking the spot where the task was interrupted, so is easier to return to it. The more time devoted to preemption, the better the performance will be when resuming the interrupted task. Research has labeled the time available for preemption of the interrupted activity as Cue Time Interval and, the time used as interruption lag (Meiran,2000)

4.3.2.F Goal activation/attention redirection

At this stage a set of rules (if present) are retrieved from memory and processing resources are assigned to the interrupting task. According to the multiple resources theory, here will be a conflict with the ongoing depending on the degree to which they are assigned the same processing resources (Wickens 2002). This model is supported by the Goal creation and activation theory described in the ACT-R cognitive architecture; where goals direct behavior (Anderson 1996). ACT-R permits activation of one goal at a time, allowing it to direct behavior. The introduction of a new goal removes attention from the active goal and focuses on the next goal. This process continues until one goal is completed or removed from the stack.

The ACT-R describes two methods for switching to a new goal: Push and Focus-on. “Push” simply pushes a new goal on top of the current one in the goal stack, even if the current goal is incomplete (see Figure 13-left). “Focus-on” pops the current goal off the goal stack, even if it is incomplete, before pushing a new one in its place (see Figure 13-right).

The Push method allows for a hierarchy of goals, with goals carried over until they are removed from the stack. Goals can be removed due to completion, outdated information, priority change, etc. When a goal is completed the next goal in the hierarchy takes its place as the active goal. The Focus-on method accounts for memory resetting and disruption, as the current goal is removed from the goal stack and returned to declarative memory in “goal chunks”. Whenever a new goal is activated, it must be recalled before it is forgotten. Anderson provides further and detailed discussion of this topic (Anderson 1996).

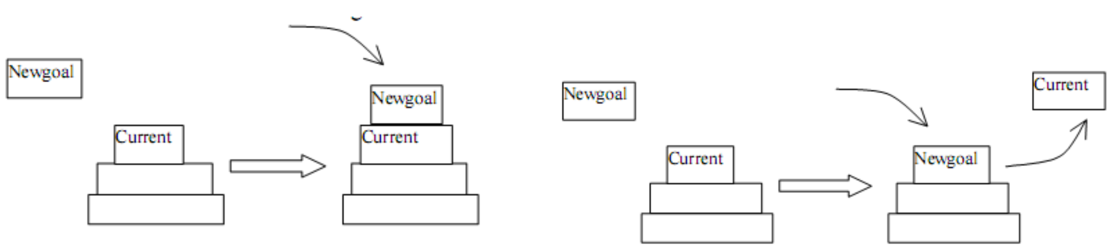


Figure 13 ACT-R goal switching models: Push method (left) and Focus-on method (right)

4.3.2.G Resumption

Completing the task associated with the interruption causes disruption, as users must transition back to the original task. While resuming, executive processes in the interruption manager query memory to retrieve the previously saved context or goal. If the context is not retrieved over time, it will suffer activation decay and will take longer to be accessed. This affects the interruption recovery time, known as resumption lag (Altmann 2000).

4.3.3 Interruption Execution Stage

Depending on the current plan and based on knowledge of how to perform specific actions for the interrupting task, a series of actions will be performed through the interface. Executive processes are also involved when the interference between two competing alternatives must be resolved, inhibiting attention to the irrelevant one. Actions can be verbally communicated messages, or physically performed procedures, such as mouse clicking, keyboard pressing, or toggle switching actions.

The flow of cognitive processing in the interruption model shows that at any given time the focus of attention can remain on the ongoing task for some time until it is captured by an interruption request to another task. The cognitive flow may also be opportunistic, when a goal-based shift is enabled by a specific state of knowledge in the problem representation (Internal interruption).

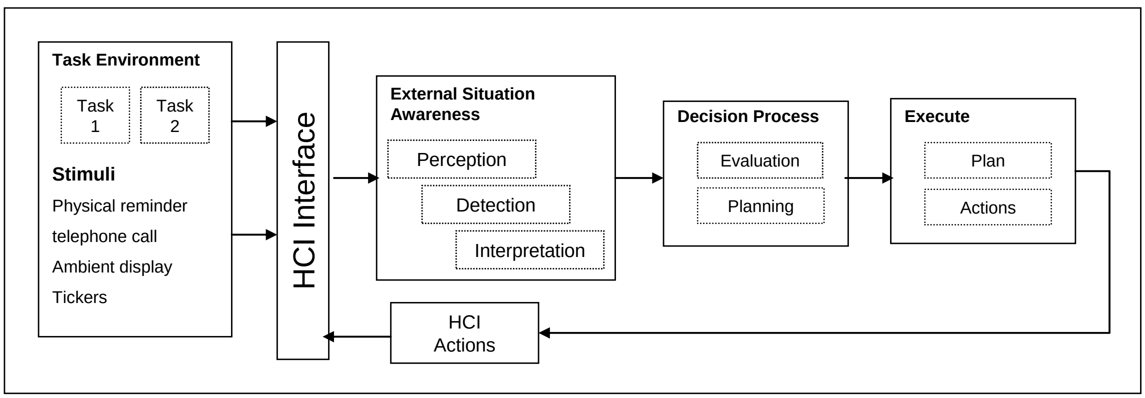


Figure 14 Interruption Process execution stage. Planned actions are performed in response an interruption.

4.4 Disruption in the interruption process

This abstract interruption model of human information processing can guide us think about the interruption process. It allows for identification of the areas prone to disruption and recognition of the parameters that might affect the interruption-disruption process.

The model presented previously identifies external awareness, decision process and resumption as three possible stages that affect the interference caused by interruptions. This interference is framed within the context of four general effects of interruptions that can be identified through the stages of the model: diversion, distraction, disturbance and disruption (Latorella 1998). Diversion can occur during detection and interpretation of an interruption Distraction and disturbance can occur during the decision process. And disruption can occur during the resumption stage.

After detection of a stimulus, people's attention is redirected away from its current focus and diverted from the ongoing task towards the interpretation of the incoming interruption. Interpretation requires attention resources in order to retrieve/activate and map the incoming interruption to representations in memory. Thus, less attention would be available for ongoing processes. This initial effect of an interruption causes diversion of resources. Once the interruption enters the decision process, further resources are necessary, requiring further attention resources to maintain a representation of the interruption in working memory. Based on the interpretation of the interruption, the decision process can disregard the interruption, act on it or schedule it for later processing. If the interruption is disregarded immediately and users return to the ongoing task, the interruption effects to this point can be considered minimum, causing small distraction on the ongoing task.

If the interruption is acted upon, the executive process must preempt information regarding the ongoing task (also add it to the competing pool of tasks), retrieve and activate a set of rules associated with the interrupting task from LTM, and form an action plan before diving into the interrupting task. If the interruption is scheduled, the executive process must add additional tasks to the pool of tasks, to keep track when the interruption will be performed (perhaps after some stage in the main task is completed) and form an action plan before returning to the ongoing task.

The process of scheduling or acting on an interruption imposes additional attention and working memory resources associated with preemption, rule activation, resumption and planning. This additional processing of interruptions can cause considerate disruption on the ongoing task.

4.4.1 Disruption in the model

This section illustrates how processes in the model can affect interference caused by interruptions. By identifying disruptive situations, it is possible to guide the design of interrupting systems.

4.4.1.A Not enough time for preemption, resumption, or rehearsal

When the interrupting activity does not provide enough time for preemption and the interrupting activity must be attended promptly, people will go directly into performing the interrupting task. This sudden redirection of mental resources will cause resources normally assigned to the executive process to be assigned to the interrupting task. Therefore, the ongoing task cannot reach a significant stop point and it will be greatly affected when resumed. In contrast, if the interrupting activity is not critical, processing of the interruption will be done at expense of the interrupting activity. Preemption will take place well into the activity associated with the interruption, interfering and possibly affecting performance. An example of this situation is when the ongoing task is switched with no prior notification; the abrupt/drastic change of the task environment becomes the associated notification to the interruption and the decision process will be activated into the interrupting task environment; which might cause preemption interference.

When the interrupting task does not provide enough time for rehearsal of the preempted interrupted task, preempted information might be lost due time decay and resumption will take longer. Research shows that the number of items that can be retained in memory decreases as word length increases (Miller; Anderson). Research also indicates that the only way information can be maintained in working memory is by rehearsal. For example, if rehearsal is prevented by an irrelevant distracting task, recall of a string of letters within the memory span decreases to close to chance after 18 seconds. (Murdock 1961, Peterson 1959).

4.4.1.B Task complexity

The level and type of resources required by the interpretation and decision processes can determine interference effects from interruptions. The higher the complexity of the ongoing or interrupting task, the more resources will be necessary to process, preempt and resume the interrupted activity (resulting in higher interference levels). As the complexity of a task increases, there is an. increased load imposed in short-term and it begins to overload. (Jacko 1997). Research also shows that complexity affects how disruptive interruptions are perceived (Gillie 1989, Bailiey 2000).

People use different strategies to form mental structures to represent ongoing and interrupted tasks. In some situations the ongoing task might be too complex to be preempted properly and these strategies might allow some detail to be lost due to task complexity. People might also use strategies that do not create good association structures, causing resumption to take longer.

It is possible for the hierarchy of tasks to have changed while executing the interrupting activity, making the interrupted task no longer important to execute. In this situation, the interruption manager has to evaluate its current goals and determine which task to execute next. There will be an associated resumption lag associated with the time required to retrieve memory constructs and goal activation before executing a new task. Another possible problem is for several tasks in the pool of tasks to have assigned high priorities, causing to the interruption manager to generate internal interruptions and multitask constantly, barely allowing time for proper preemption and resumption.

4.4.1.C Limited resources

The model assumed that the human information system has limited resources. When its limits are reached, preemption will no longer be possible or be done at the expense of existing resources/memories. The mental memory representation of the interrupted task might dissipate due to resources being allocated to the interrupting task. Interference resulting from interruptions also depends on the resources available to process them.

If resources that must be assigned to the interrupting task are similar to resources assigned to the ongoing task or to the preempted task, they might interfere with each other due to interference. On the other hand, if the set of rules from the interrupting activity are similar to the preempted task, there might be a detriment in the preempted data due to the use of similar processing resources (Sternberg 1969, Wickens 2000), but goal activation might not be affected since it might still be active in memory. Information in short-term memory is coded with three types of codes: visual, phonetic, and semantic. If interrupting activity is coded similarly to the information currently present in working memory, the resulting interference produces a large amount of forgetting. Interference accounts for the largest part of forgetting in short-memory. The forgetting effect is reduced if the interrupting activity differs from the contents in short-term memory (Wickens 2000).

4.4.1.D Familiarity and practice

When users have the ability to practice handling interrupting tasks, they become more effective at dealing and recovering from interruptions. Practice, has the potential to reduce the effects of interruptions over time(Hess and Detweiler 1994). With familiarity, constructs for retrieving frequently activated information from long term memory stores are readily available and aid in the processing of interruptions effectively. Research shows that goal activation takes longer to switch from familiar to unfamiliar tasks than switching in the opposite direction. If a novel interruption is presented, more resources will be necessary to properly interpret the incoming stimuli (might need retrieval from long term memory). People can also increase their performance in detecting interrupting stimuli, reducing diversion of resources. With practice, users might come to anticipate interruptions and preempt context ahead of time. Research shows that people perform better when anticipating interruptions that when interruption appearing randomly

4.4.1.E Recursive interruption

Recursive interruptions can be characterized as interruptions interrupting interruptions. They add to task complexity by requiring extra resources just to keep track of previously interrupted tasks. When the interruption manager schedules attending to an interrupting task a later time, but never executes it, the interruption is said to have been scheduled indefinitely. Indefinite scheduling of interruptions can be caused by, corrupted information in the pool of tasks. The pool of tasks and their associated preempted information is the most critical point for failure when recursive interruptions are present. If recursion is such that the hierarchy of goals changes constantly and/or the pool of tasks increases overwhelmingly (causing it to overflow), indefinite scheduling might be induced and the preempted tasks might be lost and very hard to recover (Miyata 1986).

4.5 Chapter Summary

The information processing model shows that depending on the type of interruption, people use different flow of attention, For example, a highly practiced interruption will be mapped from long term memory directly into Perception and the response would likely be immediate. Those memory activation and attention redirection are in part responsible for interruptions being disruptive.

The theoretical human information processing model presented in this chapter formalizes the interruption management process. This enables definition of disruption as the negative effects on ongoing tasks. The model provides a structure for formulating research questions for better understanding disruption. In addition, it serves as the base for a disruption manager that takes into consideration the mental processes depicted in the disruption model.

The cognitive models presented in this chapter aid in understanding the parameters needed for a computer-based disruption manager. They also inform the framework presented in Chapter 5, this framework is designed so that it serves as the basis for a computer-based "disruption managers". This disruption manager pre-process interruptions as humans would do it and filters them based on the interruption relevancy to the ongoing tasks and goal, among other things. From the model we determine that task and goal priority play an important role in the interruption decision process.

The model presented in this chapter guided research on the factors that influence the interruption process. Several experiments in chapter 6 evaluated factors such as, task priority and interruption relevance. These experiments explored interrupting people while performing a task with interruptions varying their relationship to the user's goal.

C H A P T E R

5

DISRUPTION MANAGEMENT FRAMEWORK

5.1 Introduction

People are constantly bombarded with interruptions from various sources; telephone calls, email, chat, and television are a few of the many distractions that people must deal with everyday. Computers related technologies have become some of the most common and highly disruptive types of interruptions. Studies have shown that various computer related interruptions such as email and instant messaging can decrease productivity and lower a person's ability to concentrate on a given task (Czerwinski 2000, Horvitz 1999). At the same time, similar studies have shown that interruptions can be beneficial, particularly when the interruptions allow a person to multitask as opposed to focus on a single task (Cutrell, et. al 2001). Since interruption is inherent to the concept of multitasking, effective multitasking requires limiting the level of disruptiveness caused by handling multiple tasks concurrently.

This chapter presents a framework designed to serve as the basis for a computer-based disruption managers. This framework outlines the factors needed to mediate disruption in computing activities regarding the interruption context and its relationships to the user's goals.

5.2 Model-guided exploration

The interruption model presented in Chapter 4 sets the framework for exploring different aspects of human disruption. The decision process in the interruption model evaluates incoming interruptions with respect to the current state of the ongoing activity or situation. It determines the priorities of each interruption to decide when to execute the task associated with the interruption. Our previous work investigating interrupting modalities has shown that the decision process is affected by people's individual differences, such as prior experiences, motivations, and psychological factors (Arroyo and Stouffs 2002). The model also identifies goal priorities, goal commitment, and goal relevancy as other factors affecting the interruption-decision process.

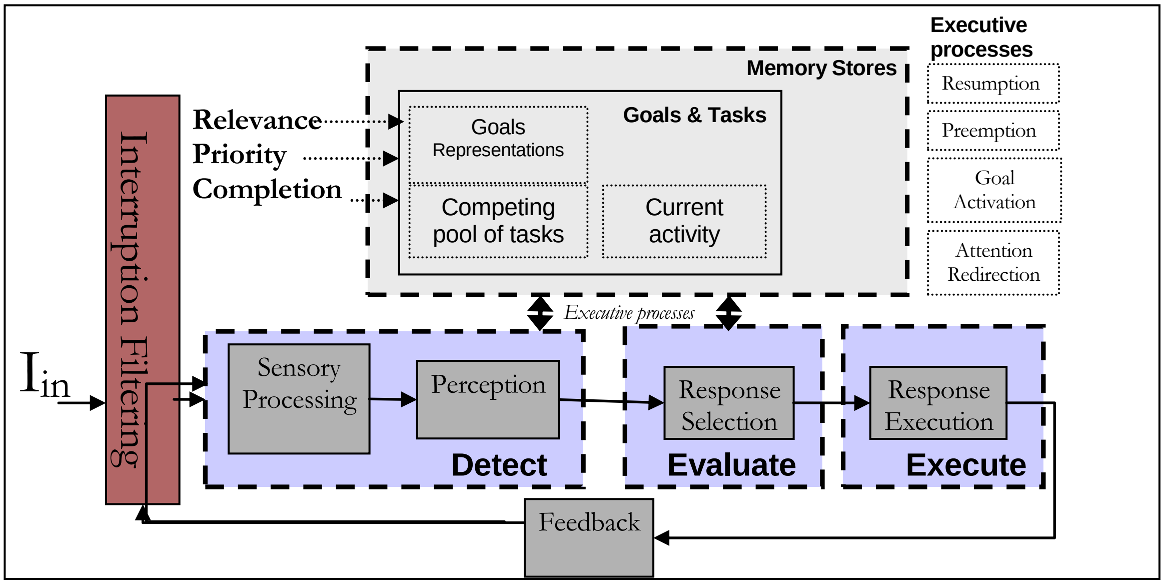


Figure 15 Based on the interruption management process, task and goal priority play an important role in the interruption decision process. Thus, it is possible to pre-filter interruptions based on task relevance and priority, as well as task completion.

One of the key questions for the understanding of human disruption is identifying the factors that play a role in people's decision process regarding interruptions. However, research has focused on identifying the factors that influence user performance in the context of interruptions. Past research has evaluated task complexity (Bailey et al. 2000), (Cutrell et al. 2001), the coordination method used to handle interruptions (McFarlane 2003), the interruption point at which interruptions arrive (Cutrell et al. 2001), the similarity between ongoing and interrupting tasks (Gillie and Broadbent 1989), the interruption modality (Latorella 1998), etc. However, the factors that influence people's decision process when evaluating interruptions have not been explored.

People react to interruptions differently depending on the circumstances surrounding the interruption and on individual factors. A factor strongly correlated to performance is the level of goal commitment, the importance of the task, and the belief that the goal can be accomplished (Locke and Latham 2002). The importance of the goal to the individuals, will affect subsequent reaction to an interruption. We explore the hypothesis that people's reactions to interruptions and perceived disruption are principally affected by goal-oriented strategies users adopt to evaluate incoming interruptions in order to accomplish their goals. As a result, user goals and motivations take precedence over pure micro-task benefits. Chapters 6 and 8 present several explorations on this subject.

We have argued that incoming interruptions are evaluated with respect to the ongoing processes goals and priorities. Therefore, people can be influenced by the level of commitment to a task. If a task is almost completed, people can opt for finishing the task before accepting an interruption and switching to the interrupting task. Our work explored the level of commitment factor and the boundaries that determine when a person is less likely to accept an interruption due to his commitment to the ongoing task/goal.

Goal and task priority play an important role in the interruption decision process. Task priorities are constantly changing as the user activities progress. But how do these changes in task priority affect people's reactions to interruptions? People can decide to accept interruptions relevant to a high priority goal or task. We evaluate the effects from interruption relevancy to the users' goal and task priority on perceived disruption. This evaluation includes exploring people's strategies for dealing with interruptions. It also investigates people's reactions to interruptions when their tasks are prioritized vs. when their tasks do not have any logical prioritization.

5.2.1 Interruption-Disruption

People treat interruptions in different ways depending on the context of the situation. Some people program themselves to ignore interruptions until they have reached a milestone in their current task. While others prefer to deal with interruptions as they arrive, in the hopes of being more productive. Czerwinski characterized the density and nature of interruptions in a diary study of task switching due to interruptions. They reported frequent and deliberate task-switching activities by typical knowledge workers over the course of a work week. Their study shows that task complexity, task duration, length of absence, number of interruptions, and task type influence the perceived difficulty of switching back to tasks (Czerwinski, Horvitz 2004). Other studies have revealed that people's willingness to handle interruptions varies across individuals with their location, as well as with their activity (Nagel, Hudson, and Abowd 2004).

Research in the area of interruptions has focused on understanding interruptions at a perceptual level (focusing on micro activities). This research has yielded important insights such as interruption relevance to the ongoing task, interruption effects at different task execution times, and so on. We support a broader study of interruptions that includes goals and implicit motivations that make people react differently to interruptions.

5.3 Disruption Mediation Framework

The disruption model and cognitive models previously presented aid in understanding the parameters needed for a computer-based disruption manager. Current interruption models do not support interactions for systems that continuously monitor the user and dynamically adjust timing and interrupting modality, among other parameters. Furthermore, interruption models focus solely on the user, disregarding the context from interrupting applications. Based on our own work and existing research we take the approach that goal concepts and task context serve as important factors in predicting disruption, see Figure 16.

The framework provides a modeling method for designing systems with interrupting goals and exposes designers to the factors related to interruption that they should consider when designing interactive systems. Our approach is meant to be usable and easily implemented under different contexts.

Several experiments have contributed to this framework. An experiment evaluated identifying what the best times to interrupt based on user activity. On this experiment we identified task switches and task completion as some of the common methods (rationale) used by subjects to decide when to interrupt someone. From this and other previous investigations, it is clear that task should be taken into consideration when mediating interruptions.

Engagement meter described in Chapter 3 tracked user activity while browsing websites and allowed people to coordinate between two activities (navigating a web site and watching a baseball video) by identifying pauses in activity and task switches. Although rudimentary, engagement meter was usable and proved the value of understanding people's activities. Further investigations in online activity tracking include systems capable of identifying people's interests and interaction types based on their mouse activity while web-browsing.

The interruption process, as described by the interruption-disruption model in Chapter 4, is closely related to attention. Attention determines which goal concepts are relevant and might also determine which tasks are important or have higher priority. The framework includes concepts to provide a cognitive representation of the user's goals and offer insight into the user's attention. .

The proposed approach allows examination of parameters from incoming interruptions in order to evaluate disruption from incoming interruptions. The framework identifies concepts, user activity, and task context as its three main components. Focusing on the requirements from interruptions themselves is a different approach from traditional approaches to the interruption problem. In our work, interruptions are evaluated regarding their context, and their relationships to the user's goals.

5.3.1 Tasks

The benefit from accepting interruptions should be balanced with respect to the ongoing task. Therefore, the challenge consists of balancing interruptions at a task level while supporting the user's goals.

Our approach extends and makes use of some constructs defined by previous research in the area of interruptions. This research indicates that disruption is related to several task factors:

1. task complexity and similarity of the primary or interrupting task (Gilli, et al 1989),
2. interruption relevancy to the primary task, task stage when interrupted (Czerwinski 2003),
3. interruption coordination method (McFralane 1999),
4. modalities of the primary task and interruption (Arroyo 2002, Latorella 1998).

Task context includes *Task level, Type of Task, Task Difficulty, Task Complexity, Task Priority, Number of Tasks, Task priority*. In addition to task context, low level factors, such as micro tasks and HCI interactions are used for mediating interruptions, as they provide a fail-safe response when no domain-specific data is available.

5.3.2 Concepts

Current approaches for reasoning about the user's goals focus on sensing user actions directed at achieving a domain-specific goal (Gievska 2004). This approach is dependent on previous examination of the desired domain and is limited to known domains or the domain itself. Instead, the framework uses concepts surrounding the user environment as a way to reason and think about the underlying user's goals. SuwatanaPongched classified interruptions into three categories (SuwatanaPongched 2005):

1. Interruptions relevant to the primary task that assist completion of the primary task.
2. Interruptions irrelevant but relevant to the primary task, although not contributing to completion of the primary task.
3. Interruptions irrelevant and irrelevant to the primary task.

These interruption categories are useful; however, people often perform more than one task at a time in order to accomplish one goal. Furthermore, several tasks can be grouped as being part of a single goal. The researcher's interruption classification schema can be extended to include the relationship to the user's goals. Interruptions irrelevant to the primary task can be related to the user's goal, thus contributing to completion of one of the user's goals. Some of these goals are unique in the sense that the user might be willing to sacrifice a certain amount of primary task attention in order to achieve them.

The model assumes that concepts surrounding the user's goals take precedence over any other factors around the interruption process. That is, interruptions might be disruptive to the ongoing task, but they might support another goal. Research agrees that irrelevant, unrelated interruptions can be harmful to the primary task, and that they elicit frustration and anxiety (Cutrell 2001, Gillie et al 1989, McFarlane 1999) . In our exploratory work and through experimentation we have also shown that people can benefit from interruptions if they are relevant to the ongoing task or the user's goals.

Existing semantic knowledgebase systems allow us to compute semantic similarity and perform query expansion from a concept. Tools such as, Cyc and ConceptNet allow for context-oriented inferences over real-world texts (Lenat 1995, Singh 2002). This type of tools makes reasoning about goals through concepts a practical approach.

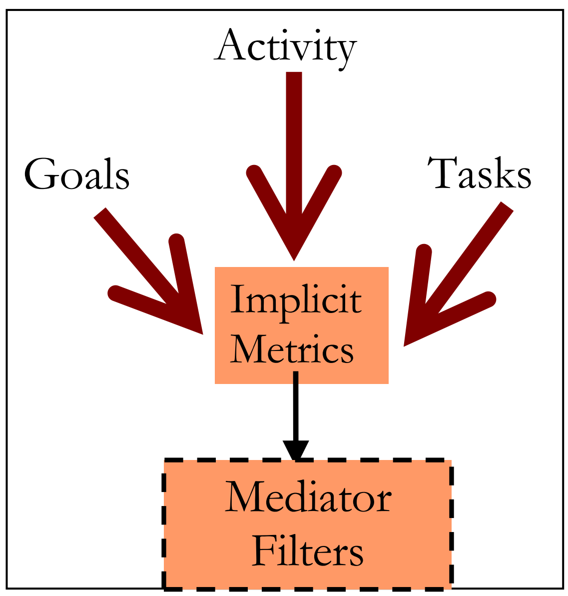


Figure 16. The framework identifies Goal concepts, user activity, and task context: the three main components required to mediate disruption from implicit metrics.

5.3.3 Activity through Virtual Sensors

The framework focuses on supporting people's goals as a mean to reduce disruption. The degree to which interruption mediated interfaces support the user goals is the key factor that determines their success. The success is also determined by inferences generated from sensors and virtual sensors. These sensors provide a narrow scope of people's goals and tasks and are limited by the accuracy of their metrics. Thus, we must identify existing sensors that generate sufficient knowledge independent from domain-specific sources. These domain-independent sources provide the basis for mediating disruption when no other data sources are available.

New tools are being developed to gather information about attention from low-level data. We developed several tools to investigate attention and user interest from domain-independent low-level metrics. In this work, we demonstrate the use of mouse tracks as domain-independent sources of information that provide relevant information about people's activities. Chapter 6 demonstrates the use of mouse data to learn about people's interests while navigating web pages and predict user activity. The use of these metrics is shown to be effective at mediating disruption when used in conjunction with task relevant data; demonstrated by a disruption manager on Chapter 7.

Sensors and virtual sensors are bound to improve; everyday new technologies and approaches provide the much needed insight into people's computing activities. However, we must consider that as new sensors are being developed, their implementation is time consuming and user acceptance increases slowly. Therefore, domain-independent sources of information are still usable as means to provide fail-safe disruption mediation.

5.4 Chapter Summary

We have presented a framework for exploring different aspects of disruption in Human Computer Interaction. The framework and interruption model identifies people's goals, interruption content, goal and task priority, and task context as factors influencing the interruption decision process. Based on this decision process, pre-evaluating interruptions based on their relevancy to ongoing goals and task is possible, associating interruption content to ongoing activities. However, it is necessary to evaluate the framework prioritization factors before creating a disruption manager. Chapter 6 evaluates and investigates if there is a main effect for prioritization, engagement level, and relevancy.

The framework's ecological validity was demonstrated and tested with a disruption manager system designed to control interruptions on web browsing and instant messaging environments, see Chapter 8. This implementation demonstrated the framework's usability with respect to practical applications.

C H A P T E R

6

DISRUPION EXPERIMENTS

6.1 Introduction

One of the key questions for understanding disruption is identifying the factors that play a role in people's decision process regarding interruptions. However, research has focused on performance in the context of interruptions. Past research has evaluated task complexity (Bailey et al. 2000), (Cutrell et al. 2001), the coordination method used to handle interruptions (McFarlane 2003), the interruption point at which interruptions arrive (Cutrell et al. 2001), the similarity between ongoing and interrupting tasks (Gillie and Broadbent 1989), the interruption modality (Latorella 1998), etc. However, the factors that influence people's decision process when evaluating interruptions have not been explored.

As described in the disruption management framework in Chapter 5, perceived disruption is assumed to take precedence over pure performance metrics. Several experiments were conducted in order to understand perceived disruption and its relationship to performance, and availability to interruptions. .

First, we evaluated the effect of interruption and goal relevancy on an ongoing task using cellular phones as part of the interrupting stimulus. An experiment evaluated the effect of relevant and irrelevant interruptions presented at times when they might have a disruptive effect on user activity. Interruptions irrelevant to the user-defined goals were perceived as highly disruptive. On the contrary, interruptions relevant to the user-defined goals were perceived as less disruptive. This behavior was observed even when under normal circumstances interruptions would be perceived as disruptive.

Second, we evaluated the effect of task/goal priority on people's willingness to accept interruptions and overall task performance. We evaluated four task/goal priority conditions: no prioritization, prioritized by time, prioritized by quality level, and prioritized by quantity. While evaluating task priority we also explored the effect of task completion at a broader time-scale (task grouped by goal, rather than tasks themselves) and its effects on people's availability to interruptions. This experiment also showed the level of commitment and boundaries that determine when a person is less likely to accept an interruption due to his commitment to the ongoing task.

6.2 Interruption Relevance Experiment

We have suggested that interruptions that take into consideration a person's motives and goals are far less disruptive than interruptions that do not consider user goals. We investigated this in an experiment presenting SMS (cell phone-based messages) interruptions while people perform activities highly susceptible to disruption. A Cell phone based software presented several SMS messages asking people to rate how disruptive the message was to their ongoing activity—paying attention to a speaker on a video. Subjects rated how receptive they were to the incoming message and if the incoming interruption was perceived as disruptive. Incoming interrupting messages had different timing (before task, during task, and never) randomly presented trough each of the video sessions, and varied content (from relevant to irrelevant); some SMS were irrelevant to the user task, such as random jokes, or quotes.

Our hypothesis stated that interruptions irrelevant to the user-defined goals would be perceived as highly disruptive. On the contrary, interruptions relevant to the user-defined goals would not be perceived as disruptive.

6.2.1 Experimental Platform

The test platform for cell phones presented in this section goes beyond the traditional approach of studying interruptions by using computers. The platform provides the ability to conduct additional experiments to test the effect of interruptions during a wider range of activities that don not necessarily involve computer related tasks (more uncontrolled situations than the desktop environment). The cell phone offers a way to study interruption in new contexts and can validate and build upon the findings of previous studies. .

The advantage of using cell phones is that it tests disruption effects on natural settings, users may be interrupted during normal routine activities such as driving, walking, working, etc. Furthermore, cell phones in their very essence provide an ideal means by which to interrupt users. People have become used to being interrupted by their cell phones at unexpected moments. People are also accustomed to carrying their cell phone around with them throughout the day so designing an experiment that uses a cell phone does not introduce an unfamiliar scenario. Lastly, modern cell phones posses an incredible amount of processing power and features that can be used to detect user context. These features lend themselves well to the concept of studying interruption because a system can potentially be developed that uses the various sensors on a cell phone to identify the user's context and possibly interrupt them under specific conditions. The different features that make this possible include Bluetooth, GPS, and Internet connectivity.

6.2.1.A System Description

This section presents the design of a system that monitors the level of disruption caused by interrupting people during various activities. The system is based on a cell phone application that periodically interrupts users with incoming messages intended to divert their attention from an ongoing activity. Interruptions are controlled by an experimenter who may choose to interrupt the users during specific activities. The application logs data about the users' context as well as information about the users' responses to the messages (including the level of disruption caused by the interruption). The application can be used to conduct a wide variety of experiments to study human response to interruption.

A cell phone interruption application was developed as an experimentation platform. It is a client-server application coded in Java that uses J2ME standard for mobile devices and a server Java Servlet API and running on an Apache Tomcat server.

6.2.1.A.1 Client

Upon execution, the application allows the experimenter to select one of six configuration modes; one experimenter mode and five other user modes. The experimenter mode allows the experimenter to control the conditions by changing the activity to be interrupted, which controls which cell phones will receive an interruption. In user mode, a survey screen prompts the user to input identification information. Figure 17 shows the screen used to get an idea of how accustomed user are to using cell phones as their response time is measured and logged in permanent storage when finished. This data (along with the other experiment data) may be uploaded to the server at the end of the experiment.

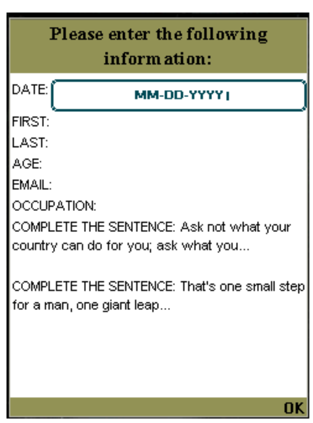


Figure 17 Survey Screen provides information about the user's ability to interact with cell phones.

When the experimenter changes the condition, all cell phones with a configuration mode that matches the new condition, receive a series of interruption messages signaled by an audible alarm and vibration; see Figure 18-left. A screen with more information about the interruption only appears if the user chooses to accept the incoming message; see Figure 18-center. During an interruption, the user response times and response values are stored in a record store log on the phone. The application continues to interrupt the participants until either the experimenter ends the experiment or the message database on the phone runs out of interrupting messages.

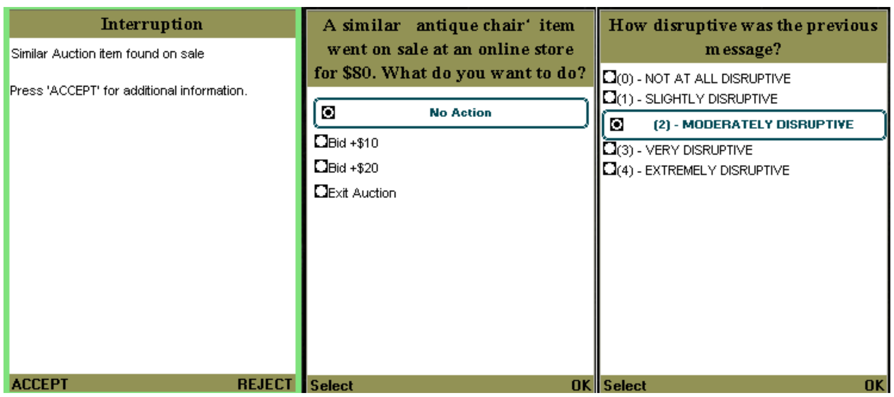


Figure 18 Sample of relevant SMS notification (left), SMS message content and available options (center), and perceived disruption survey (right).

6.2.1.A.2 Server

The server application receives GET and POST requests from the experimenter cell phone to update the experimental conditions, while client cell phones poll the server periodically to determine the current experiment condition. The server is also used to store interruption messages, so that they can easily be changed to address different experimental questions. The messages can be downloaded to each cell phone by opening a stream to the server and storing each message into a record in a message database on the cell phone itself. Similarly, the server is used to receive and store experiment data. This allows updating the rules in each condition that determine when interruptions should be presented.

6.2.2 Experiment Description

A pilot study was conducted using the cell phone application to gather preliminary data about disruption caused by interrupting people with SMS messages relevant to a secondary goal while busy on a primary task.

	Avg. Time to Accept/Reject	Avg. Time to Select an Option	Avg. Time to Rate Disruption Level	Avg. Disruption Level
User 1	2.1 (2.2) sec	10.4 (n/a) sec	2.4 (14.1) sec	2.0 (4.0) / 4
User 2	9.5 (8.6) sec	15.1 (n/a) sec	6.6 (9.4) sec	3.0 (5.5) / 4
User 3	3.7 (1.9) sec	7.5 (n/a) sec	2.3 (39.4) sec	1.7 (5.0) / 4
User 4	5.1 (5.4) sec	11.8 (n/a) sec	7.1 (18.0) sec	2.3 (3.5) / 4
User 5	4.5 (2.4) sec	8.7 (n/a) sec	3.5 (14.3) sec	2.0 (4.0) / 4

Table 1 Average results for each user (values in parenthesis are for irrelevant messages)

The pilot study with five subjects indicated a gradual decrease in average response time between the first and fourth interruption, see Table 1 for detailed timing information. It is very likely that this downward trend in response time was simply due to the subjects becoming more accustomed to the cell phone interface. This indicated that subjects became proficient with the interface and interrupting scenario after the fifth interruptions; suggesting that subjects needed more practice time. The pilot study showed that valuable data can be obtained from experiments designed around the basic application framework. It is important to realize that changes in the experiment conditions were possible using the framework of the interruption application.

6.2.2.A Protocol

The experiment began with the experimenter handing out the phones pre-loaded with different interruption configuration modes. Twenty subjects completed a pre-experiment questionnaire and they all reported that they had substantial experience using cell phones. The number of participants per session varied between 2 and 4 subjects per session. A cell phone-based multiple-option questionnaire tested participants' familiarity and dexterity using cellular phones. This pre-experiment procedure was also designed to get people familiar with the cell phone interface. Participants were told to answer the survey questions presented on the phone. Once each participant signaled that they were finished with the survey, the experimenter gave them the following oral instructions:

> *"You are interested in bidding on an antique chair at an online auction. The cell phone that you have is your portal to the auction. It will periodically, update you on the status of the auction and ask for your input. You may choose to respond to the updates or you may ignore them. At the same time, your primary task is to watch and listen to the video presentations of a series of professors as they explain their research. It is important that you pay attention to the presentations because following the video you will be tested on how much you recall from the presentations."*

Each participant was given a Nokia 6630 cell phone loaded with the interruption application described in the experimental platform section. They were also given headphones and presented with a projected video consisting of a series of brief presentations from a conference. The use of headphones kept distractions to a minimum and kept subjects from interacting with each other.

Participants received 6 test interruptions in order to familiarize themselves with the cell phone interface and identify any potential problems. Incoming messages were signaled by a soft chimes sound gradually increasing in volume for about 30 seconds. Once finished, participants responded to the test interruptions and asked any remaining questions about the procedure.

Participants were then told to put on their headphones and the experimenter started a video presentation that lasted approximately 30 minutes. The video included 6 speakers presenting their research work for approximately 5 minutes.

At the experimenter's discretion, the subjects were then interrupted with messages throughout the video. The experimenter used a cell phone to control the flow of messages sent to the participants. The message content was "relevant" content, meaning that it was relevant to their secondary task of bidding on an antique chair. Occasionally however, the content was "irrelevant" content which had nothing to do with the user's goals. These messages were simply random news headlines of no relevance to their user. The session included 16 messages; 50% irrelevant and 50% relevant.

At the end of the video, participants were given two questionnaires. The first questionnaire asked the subject to list as many of the research groups and professors that they could remember and provide a brief summary of their research (based on the content of the video). The second questionnaire presented the subjects with two lists; one list containing the names of the professors in the video, and the second containing the names of research groups. The lists were lined up incorrectly and the participants were told to do their best to match each professor with their corresponding research group. After completing the questionnaires, subjects were asked to comment verbally on the experiment and the overall disruptiveness of the messages. Following their comments, the subjects were released.

6.2.2.B Results

The results show that that, on average, irrelevant interruptions were more disruptive to participants than relevant interruptions. On average, subjects took 6.7 seconds to acknowledge each relevant message while they took 14 seconds to acknowledge irrelevant messages. This indicates that irrelevant messages were indeed irrelevant and were place on hold until subjects felt it was appropriate to reply (or disregard them). In complex situations people are generally good at identifying useful and relevant information.

The task was successful in engaging subjects as some of them demanded to know the outcome of the online auction, and whether or not their actions were successful. On the other hand, the primary task proved too delicate, as subjects often missed important information when interrupted; resulting on a 33% average score across all subjects. After filtering for outliers, there was no noticeable downward trend in the average disruption level reported by the users on each condition (disruption levels for relevant or irrelevant messages remained constant through the experiment). This indicates that any differences in disruption levels were accounted by whether or not the message was useful to the users' goals.

Furthermore, perceived disruptiveness, as measured by the user's selection of a disruption level (1-5 scale), was higher on average for irrelevant messages (4.32) than for relevant messages (2.12). These results are summarized in Figure 19. The results show that random interruptions, that are irrelevant to users' current goals and activities, are significantly more disruptive than relevant interruptions. The results suggest that relevant interruptions, even though responded-to promptly, were not perceived as disruptive due to their perceived benefit in helping subjects accomplish their goal. These experiments confirmed that people accept interruptions that are relevant to the ongoing goal faster than interruptions from irrelevant goals and than relevant interruptions are not perceived as disruptive.

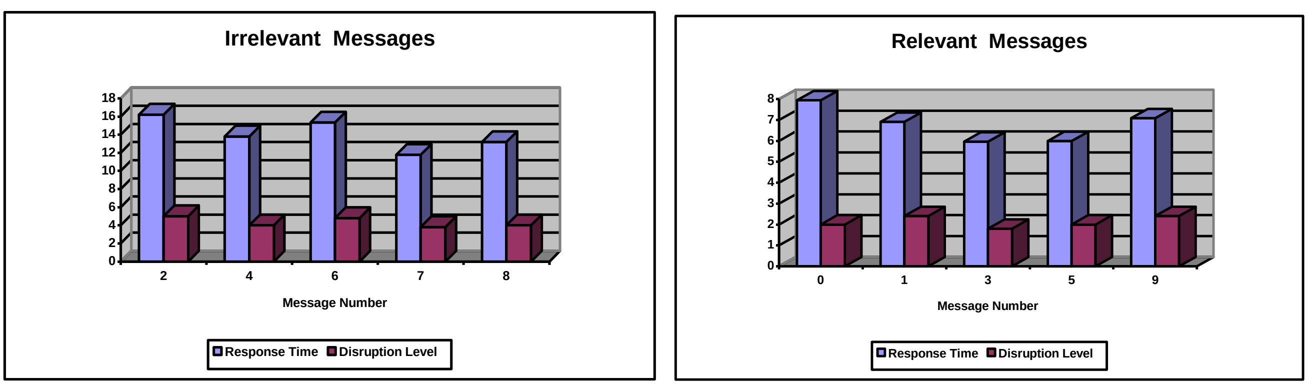


Figure 19 Average response time and perceived disruptions for SMS messages. Random irrelevant interruptions, are significantly more disruptive than relevant interruptions

6.3 Experimental Environment for Disruption Studies

An experimental task was developed to study disruption and validate a disruption mediator on a realistic scenario. The task was constructed so that it resembles typical multi-tasking activities performed on desktop computers.

In the workplace, many people use computers to perform general purpose tasks, such as word processing, Internet browsing, e-mail and other digital messaging, multimedia, computer programming, etc. A 2003 computer use survey from the Bureau of Labor Statistics of the U.S shows that 55.5% of the total (77 million) work force in the U.S. uses computers and the internet at work (DeBell 2006). A more recent survey estimates that 694 million people, age 15+, used the Internet worldwide in one single month (ComScore 2006). These common computer activities were included in the experimental task, requiring subjects to switch between word processing, reading and responding to email, Web browsing, and performing simple mathematical calculations.

Investigations in the nature of desktop interactions indicate that task switching and interruptions are typical and often deliberate actions of knowledge workers. Studies also indicate that task complexity, task duration, and task type influence the perceived difficulty of task-switching in typical environments (Czerwinski, Horvitz 2004, Jackson 2001). In an effort to use a realistic experimental task, a complex scenario with varying task traits was used. The task differed from traditional laboratory-only type tasks in that it was placed around a realistic scenario so that subjects could easily understand it and play the appropriate role for the situation.

6.3.1 Experimental Task and Scenario

The scenario consisted of customer service and order processing activity for an e-commerce site. The scenario reflects the fact that the most commonly reported task for 77 million workers who used a computer at work was accessing the Internet or using e-mail. The scenario described a typical small business environment where customer service representatives take email orders from several customers a process each order individually trying to satisfy the customer's demands and complete a sale. Small to medium e-commerce sites still require human intervention and process customer orders in a similar way.

The scenario explained that customer service representatives obtain a commission based on their sales and instructed subjects to play the role of a customer representative. Adding this role guaranteed that subjects would perform the task to the best of their abilities and encouraged subjects to obtain a bigger profit.

The customer service scenario serves as a good proxy for representing high level (abstract) goals. Rather than specifying detailed task-based goals, the scenario presents an abstract goal requiring subjects to create their own definition of the goal based on the task constrains. The abstract task goal is to satisfy the customer demands based on their requirements. This type of setting addresses the situation where a system might not be able to identify the exact representation of the user's goals, and relies instead on the specific concepts surrounding the user's goals.

The scenario portrayed the idea of evaluating algorithms used by a system designed to filter customer e-mails and introduced an Instant Message (IM) System for price updates, and requests from fellow employees. Centering the experiment on the email manager system removed subjects' attention from the real experiment (their own reactions to IM) and guaranteed that people would react to interruptions as they would normally do.

6.3.1.A Summarized Scenario:

The main concepts presented in the introductory scenario are the use of a customer service task. The task is part of the work done at a company, the main task is processing email orders, the task includes different types of customers, and the importance of satisfying customer's requests.

> *You work for a large supply company in the purchasing department. The company is testing a system that assigns different type of customers to different employees through the day, as to maintain a balanced workload. The company has also implemented an Instant Messaging system that allows its employees to share pricing information with one another. Sharing information benefits the company and you might receive a bonus based on the company performance.*
>
> *The new system classifies and sorts customer emails depending on the type of service requested and the customer's demands.*
>
> - *Customers demanding high accuracy levels, up-to-date prices, error-free orders and a high quality of service.*
> - *Customers demanding their orders to be processed as fast as possible. Timely processing is their number one priority.*
> - *Customers from medium-big companies interested in meeting purchasing quotas. They want all of their orders processed.*
> - *Low volume customers with no accuracy, time or quota restrictions. These customers have no specific demands.*

As part of your job, you receive many requests from multiple customers over email regarding products they would like to buy and get a price quote. Your task is to find the online catalog price for each of the items requested by the customer and update a customer order. Your job also includes making decisions so that customers are able to buy as many items as possible while accommodating their preferred products; all within their requirements. Customers rely on your information and intuition and will ultimately place an order based on your suggestions, however, it is in the best interest of the company to keep your customers satisfied. Customers with their demands met improve the company's economic performance.

6.3.1.B Task Description

Each order included a short email script where customers described which items they wanted to buy and why. The short email script included enough information to convey the expectations and motivations from real clients and hint the condition type. Each email script was designed so that it would remind subjects how they should process the email and to reduce the task completion time. The following email script illustrates a customer with specific demands and that should be processed with extreme care:

Hi,

We are an interior design firm expanding to new locations and are looking for a new provider to handle large Crate and Barrel purchases. Please provide competitive quotes for the following products available at the Create and Barrel Catalog. We will inform you of our decision based on your quote.
Budget: $1,450

10 Document Frame
5 Kyoto Lamp
9 Loft Three-Shelf Cart

Tiffany Graff
Design Acquisitions rep.

Subjects worked with an email client that sorted customer emails and placed them in separate folders. Each of these folders had to be completed before moving to the next one. The task required subjects to scan their email folders, decide which type of customers they might be working with and what type of service provide to these customers. The task required subjects find the listed price for the items requested by each customer (one customer at a time), and arrange the products so that the customer was able to buy as many items as possible while accommodating their preferred products; all within the customer's budget. Subjects were encouraged to use their own intuition and taste in order to get them involved in the task.

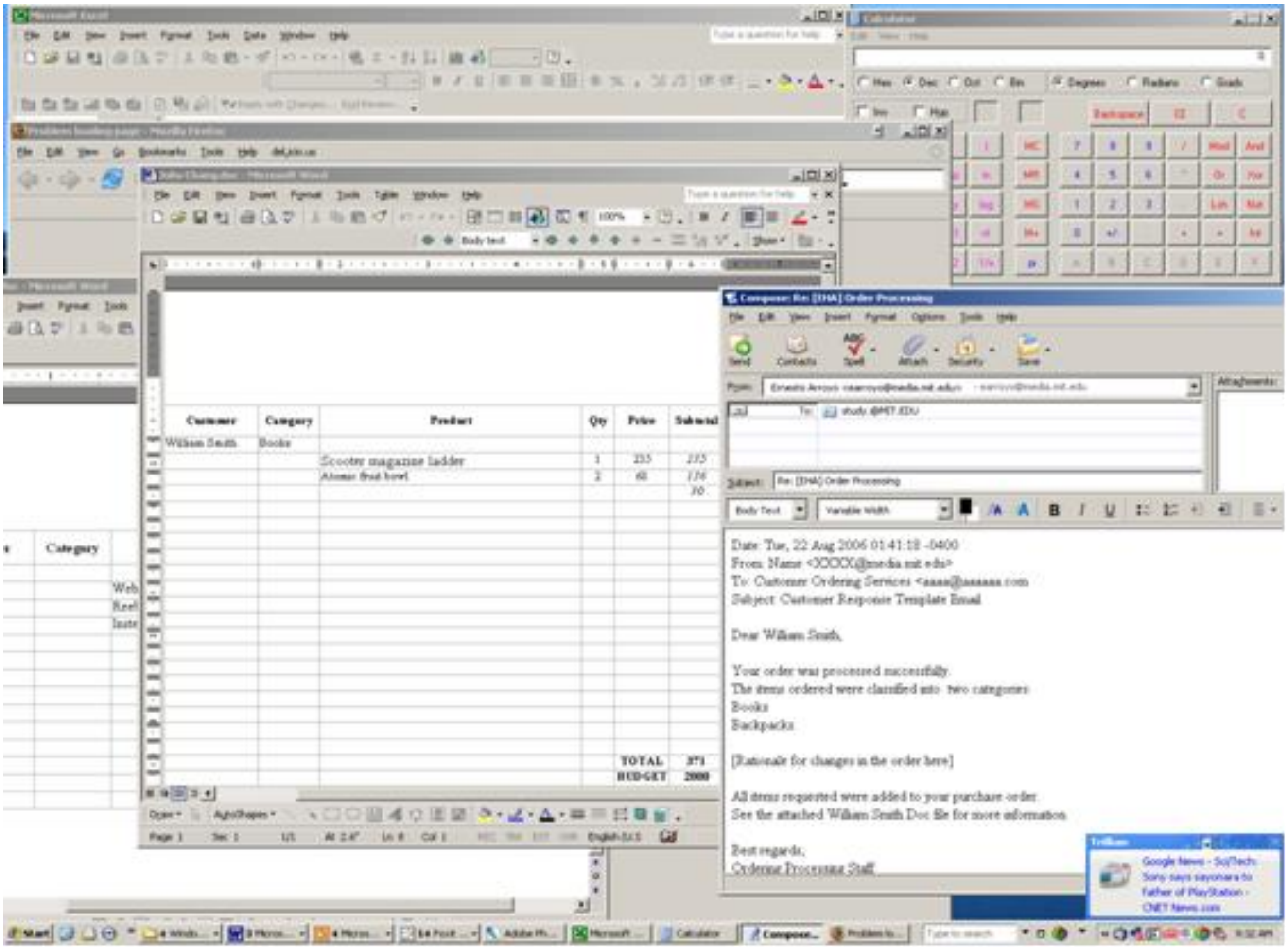


Figure 20 Experimental Task Interface. Multitasking environments require users to quickly switch contexts and make decisions that might cause high cognitive loads.

The experiment required subjects to read and process customer's emails one by one, before moving to the next email. Each email order specified which catalog should be used for pricing information; these catalogs were available online. Subjects kept track of all of the items by entering pricing information into a word template subdivided (see Appendix D.1). The template had already been pre-filled with order details and subdivided by product category so that subjects would focus on updating pricing information rather than get distracted with formatting details. This also helped reduce the time necessary to process an email order. Subjects calculated subtotals per category, grand total and difference from the customer's budget using calculator software provided to perform calculations. Once finished, subjects emailed the customer and included some of the order details, as well as any modifications to the original order. Figure 20 shows the different applications used during the experiment. The email included a summary of the categories created, total, and a brief explanation regarding unavailable items. See email template in Appendix D.1

Two out of four emails had artificially introduced budget errors. Budget errors required subjects to drop one or two items from the order, correct the item price, and recalculate the order quantities based on the customer budget restrictions. Emails with no artificially introduced errors received prices updates so that the update could be reflected on the email price. These artificial errors were used as a metric to ensure subjects devoted enough attention to the task and to keep the task from becoming monotonous. The number of items per email order varied from 3-5 items to keep the scenario more realistic, short enough not to overwhelm subjects, and to provide enough time for interruptions in the middle of the task.

6.3.2 Pilot Studies

Several pilot studies evaluated the parameters needed for the experimental task and scenario. Length of each condition was reduced to 10-15 minutes long. The number of emails to be processed by condition was reduced from an earlier experiment with 6 emails to 4 emails so that subjects wouldn't be overwhelmed with too many orders needing processing.

The pilot studies revealed subtle differences in processing/completing the experimental task based on subject's computer experience level. Rather than controlling for this effect, we decided to explore this effect further and evaluated subjects' computer experience using a self-assessment test and a computer familiarity questionnaire. The number of interruptions was reduced from 5 interruptions per email to 3 in order to allow subjects to successfully complete the order without excessive interruptions. The pilot indicated that subjects should perform a minimum of two email orders before becoming familiar with the task (Hess 1994).

6.4 Task priority effects experiment

The experiment presented in this section evaluates task prioritization effects on people's susceptibility/availability to interruptions. It investigates how disruptive interruptions are perceived depending on the type of task prioritization. The experiment evaluates four task prioritization levels: Time, Quality, Quantity, and No prioritization.

Task prioritized by time are deemed urgent or are tied to a deadline: *Customers demanding their orders to be processed as fast as possible.* Tasks prioritized by quality should be performed with extreme care and attention to detail is important: *Customers demanding high accuracy levels, up-to-date prices, error-free orders and a high quality of service.* Tasks prioritized by quantity require many small repetitive tasks to be completed: *Customers from medium-big companies interested in having all of their orders processed.* No prioritization refers to the approaches taken by people when no obvious prioritization information is available: *Low volume customers with no accuracy, time or quota restrictions. These customers have no specific demands.*

6.4.1 Experimental Design

The experiment measured subjects' availability to interruptions and perceived disruption on four task priority conditions: time, quantity, quality, and no pressure. The experiment used a repeated-measures design with the number of accepted interruptions, and reported perceived disruptions as dependent variables.

An enhanced disruption manager with logging capabilities recorded usage data. The experimental software included two modules; an interrupting module and a data logging module. The interrupting module delivered pre-defined IM messages at appropriate times depending on the condition being tested. The software module varied the type, relevancy and timing of IM presented to subjects. The logger module recorded relevant data about subjects' reactions to interruptions: mouse activity, window switches, active window, time spent on each window, type of window, IM activity, IM acknowledge time, time spent on IM, IM response, interrupted window, active window after interruption, whether or not the IM window was left unattended, among other low level and experimental control information.

6.4.1.A Hypotheses

Goal and task priority play an important role in the interruption decision process. Task priorities are constantly changing as the user activities progress. But, how do these changes in task priority affect people's reactions to interruptions? People can decide to accept interruptions that are relevant to a high priority goal or task.

Our hypothesis states that there is a difference in the way people react to interruptions depending on the task prioritization level. In other words, perceived disruption and availability to interruptions depends on the type of task/goal prioritization. Prioritized tasks (PT) are those relevant to the user goals and necessary to accomplish those goals.

Interruptions during highly prioritized tasks will be perceived as less disruptive than interruptions during non prioritized interruptions. Tasks prioritized by Time (urgency) and Quality will be perceived as less disruptive than non-prioritized tasks.

Availability to interruptions will be lower for prioritized tasks than for non-prioritized tasks. Tasks prioritized by Time, Quality and Quantity will be less likely to accept interruptions than non-prioritized tasks

6.4.1.B Exploratory questions

The experiment evaluated which type of interruptions people pay more attention to, the effect of task completion level, interruption timing, task level, and interruption task relevancy. We investigated how long people take to stop accepting irrelevant interruptions, and whether this is based on the prioritization level. People might create a mental model that includes task priority. This model can be reflected by the type of interruptions that people decide to attend to.

6.4.1.C Participants

36 subjects (21 male, 15 female) were recruited and compensated for participating in the experiment. Subjects were a mix of undergraduates and graduate students, as well as office employees between the ages of 18 and 48. Subjects had various level of computer experience, with at least one year of computer experience, Appendix D.3 includes the user experience questionnaire used. 4 subjects were removed from the analysis due to unsuccessful task completion or due to errors in data collection.

6.4.1.D Protocol

The experiment included an introductory session describing a compelling scenario where subjects were instructed that they would be employees from a company that processes sales orders. All subjects were presented with a general overview of the scenario. The scenario implied the task and the possible conditions in the experiment. The task description included details about how orders should be processed and the steps that should be taken when processing an order and where to find information; see Chapter 6 and Appendix D.1 for detailed scenario information.

Once subjects became familiar with the experiment scenario, they received a detailed task description that reinforced the concepts initially introduced. The experimenter also demonstrated how to complete the task step by step (processing an order). Subjects later performed a training session to familiarize themselves with the experimental task and the applications used. Subjects were allowed to arrange the applications according to their preferences. At this stage, subjects were presented with cases from all possible conditions as to avoid bias due to practice over one single condition. Each of the conditions was randomized to account for presentation order learning effects. Practice on all possible cases guaranteed that all subjects were exposed to the same type of stimulus. Instructions were available at all times and incentives were offered, making subjects believe they would get a sale commission if orders were processed promptly and properly.

Subjects received several interruptions (Instant Messages) throughout the duration of the experiment. Subjects would decide whether or not to pay attention to IM and respond to them, or simply ignore them. Subjects received IM at different stages within the task: reading, responding to email, browsing, reading online catalog, entering pricing information, editing text, or performing mathematical calculations.

In order to keep subjects from expecting interruptions, the number of interruptions per condition was randomized. Subjects received a maximum of three interruptions per email. Interruptions could either be relevant or irrelevant to the ongoing email: two irrelevant and one relevant interruption or two related and one irrelevant interruption. Interruptions included price requests or prices updates. Updates were related to emails with no budgeting errors. The type of interruption, order and timing of interruptions varied randomly while subjects processed each email. The type of interruption also varied depending on the type of email being processed so that each email only had one error needing attention.

Interruptions could be processed in several ways, which are reflected by subjects' reactions to the interrupting Instant Message. Subjects could either ignore or acknowledge the IM and either act upon it or take no action:

- Ignored
- Acknowledged –No action taken
    - Closed IM window quickly
    - Left IM window unattended
- Acknowledged –Action taken
    - Responded to IM
    - Switch activity based on IM content

Subjects were able to acknowledge an IM alert by clicking on it; the IM alert window would fade away after 10 seconds if not attended to. Clicking on the IM alert window would open an IM client with the IM content and subjects could either respond or close the IM window; the IM window would fade away by itself after 50 seconds of inactivity.

Each subject completed 4 counterbalanced conditions: one condition per each priority type (no priority, medium volume, urgent, quality). Each condition included 4 emails processed in approximately 15-25 minutes. Subjects responded to a short questionnaire after each condition. This questionnaire was designed to assess perceived disruption and workload—modeled after the NASA-TLX workload assessment, see Appendix D.2. All four conditions were completed in a 1:30 – 2 hours session.

6.4.1.E Results

Outliers and cases with missing data were removed from the analysis. The system logs were correlated with IM logs and later filtered in order to obtain accurate IM-acknowledge ratios per condition. Performance score was calculated based on accuracy and processing speed for each of the conditions. It represents how well the order matched customer demands. Finally, reported perceived disruption was standardized for all subjects.

6.4.1.E.1 Control hypothesis

Performance metrics indicate whether subjects processed the tasks on each priority condition according to the experiment guidelines. Performance reflects different subjects' approaches under each priority condition; see Figure 21. The neutral condition—with no priority defined, yielded the lowest score (68.5%) as no expectations were set. On the other hand and as expected, the Quality condition demanding attention to detail yielded the highest performance score (82.8%). The Quantity and Urgency conditions yielded similar performance scores (77.1%) as they have communalities in the way subjects approached them. A repeated-measures ANOVA showed no significant priority by experience level interaction and a main effect for priority type on performance $F(3, 69) = 3.768$, $p=.01$. Within-subjects contrasts show significant (at the .05 confidence level) differences between the neutral condition and the baseline or no priority condition.

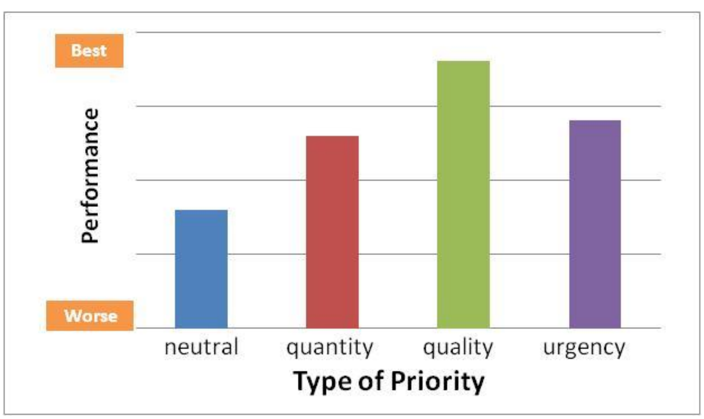


Figure 21 Performance confirms that subjects used different strategies under each priority condition in accordance to the experimental scenario.

6.4.1.E.2 Perceived disruption hypothesis

Differences in perceived disruption are shown in Figure 22. Subjects perceived incoming interruptions when working on tasks prioritized as urgent or quality as less disruptive than on the other conditions (3.6-3.7 vs. 4.4-4.8). One-way repeated- measures ANOVA with computer experience as a blocking variable indicated no significant interactions and a main effect of priority type on perceived disruption $F(3,75)= 2.9$ $p=.036$; confirming our first hypothesis that interruptions during highly prioritized tasks would be perceived as less disruptive than interruptions during non prioritized interruptions. Contrasts among the priority conditions indicate that Quality and Urgency are significantly different than the Quantity and Neutral conditions $F(1,25) = 5.3$, $p= .03$, $.05$ respectively. The Quality and Urgency conditions can be considered highly prioritized as confirmed by similar perceived disruption levels $F(1,25)= .96$, $p=.61$.

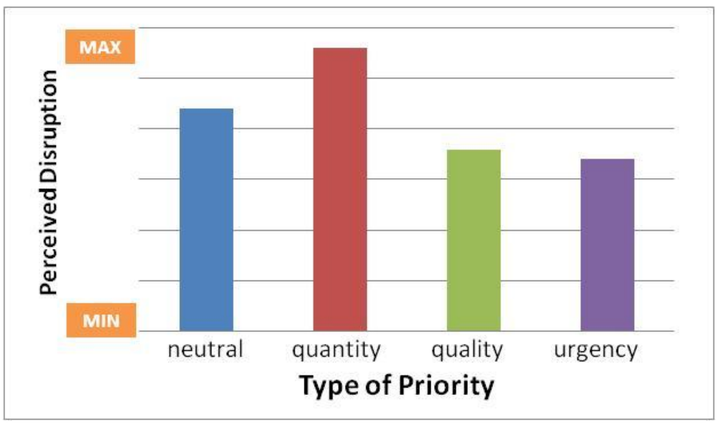


Figure 22 Perceived disruption based on task priority. Relevant interruptions during highly prioritized tasks were not perceived as disruptive.

Subjects evaluated the potential benefits from being interrupted and adjusted their susceptibility to interruptions, as interruptions could provide helpful information to complete their task. This explains why instant messages in these conditions were not perceived as disruptive, even though the number of irrelevant IMs remained the same for all conditions.

The Quantity condition presented the higher perceived disruption level. This condition was designed to elicit the sense that the task was extremely long. Thus, if we take into account that subjects felt the task would take a long time, and that it required their focus and concentration to finish them; then, interruptions that could detract them from their goal wouldn't be welcomed. This explains why subjects perceived interruptions during this task as highly disruptive.

6.4.1.E.3 Availability to interruption hypothesis

Computer user experience accounted for high variance in the number of instant messages acknowledged (IM) and the IM acknowledge ratio. Advanced users reflected a consistent behavior, with a high IM acknowledge rate under the no priority condition, a higher IM acknowledge rate on the quality condition, and a lower than neutral IM acknowledge rate under the urgent condition. On the other hand, beginners acknowledged IM differently. They acknowledged most of the IM received on the neutral condition (baseline behavior), lowered their IM acknowledge rate on the quantity and quality conditions, and increase their acknowledge rate on the urgent condition. Low acknowledge rate on tasks prioritized by quality and quantity can be explained by subjects focusing on the task at hand.

IM acknowledge rate behavior by beginner and expert users support the idea that expert users were able to better cope with Instant Messages. Advanced users acknowledged less IMs than beginner users whom, on average, acknowledged most of the received IMs. This was reflected on the neutral and urgent conditions. Figure 23 shows IM acknowledge ratios for beginner, intermediate and advanced subjects under the four priority conditions.

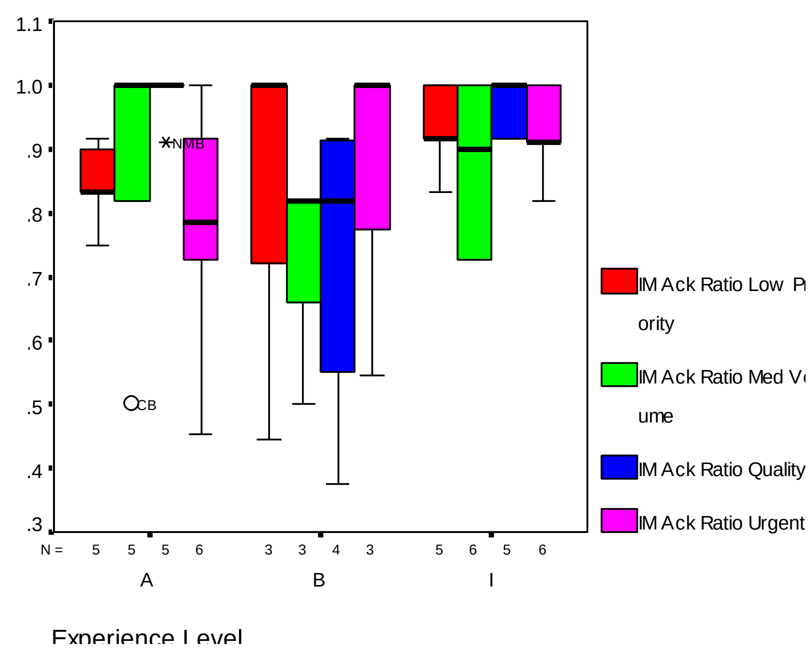


Figure 23 IM acknowledge rate by Experience level and priority condition. On the neutral condition (Low Priority), advanced users acknowledged less IMs than beginner users whom on average acknowledged most of the IMs received.. This is also reflected on the urgent condition, as expert users lowered their acknowledge ratio, while beginners acknowledged most IMs. *A= Advanced, B = Beginner, I=Intermediate*

6.5 Chapter Summary and Discussion

This chapter evaluated the effects from interruption relevance to the users' goal and task priority on perceived disruption. A series of experiments explored people's strategies for dealing with interruptions while varying the task priority level. These experiments used a configurable interruption system to test how people evaluate incoming interruptions when dealing with multiple tasks and priorities.

We have argued that incoming interruptions are evaluated with respect to the ongoing goals and priorities. Therefore, it is possible for people to be influenced by the level of commitment to a task. If a task is almost completed, people can opt for finishing the task before accepting an interruption and switching to the interrupting task. The work explored the level of commitment factor and the boundaries that determine when a person is less likely to accept an interruption due to his/her commitment to the ongoing task.

General consensus indicates that instant messages can be disruptive. That is, people use instant messages as a way to stay informed and have access to information that might be relevant to their goals (whether personal or work related). Our investigations confirmed that people tend to accept most instant messages and used this effect to evaluate how varying goal priority would make interruptions (instant messages) be processed and perceived differently. We demonstrated that goal/task priority influences people's susceptibility to interruptions. Potentially relevant interruptions are perceived as less disruptive when the user priorities demand attention to detail (quality) or are time-restricted.

This chapter, together with Chapter 4 described some of the key process of people being interrupted in Human Computer Interaction. This chapter provided some insight about how interruptions should be presented depending on their relationship to the user's goals and the task/goal priority level, overall task completion stage, and computer experience. This chapter also demonstrated that interruptions are evaluated with respect to ongoing processes goals and priorities. Therefore, there is a greater value from an interface that monitors and controls interruption with the goal of decreasing perceived disruption rather than just focusing on performance. This can be accomplished by using our findings to mediate interruptions based on goals and priorities. However, the challenge is attaining goal-related and priority information. The next chapter explores and validates using low-level implicit metrics to detect user goals and behaviors. A disruption manager uses domain-independent metrics to draw inferences and extract concepts relevant to the user goals. The manager uses text and mouse tracks low-level implicit metrics/inputs for mediating disruption.

C H A P T E R

7

VIRTUAL SENSORS AS IMPLICIT METRICS OF ATTENTION

7.1 Introduction

Collection of implicit measures of human actions is an efficient alternative to collecting or gathering explicit user feedback, which can be costly in time and resources. Implicit behavior detection is used to find out indications of user interest and ongoing activities. Implicit metrics refers to events in computer interfaces that can be automatically captured to analyze and predict user-related topics and interests.

This chapter demonstrates that concepts surrounding the user activities and mouse movement are reliable and accurate indicators (implicit metrics) to control user disruption. Documents and mouse tracks provide insight onto the user goals, priorities and behaviors. Domain-independent text streams are used to draw inferences and extract goal-related concepts, while mouse movements illustrate hidden behaviors reflecting user's attention.

A disruption manager specifically designed for web-browsing can use mouse tracking data to infer browsing behavior, extract IM relevant concept topics and mediate incoming interruptions. The manager can delay interruptions based on their relevancy to the user's goal topics and current user activity. If the user is actively reading an online article, the manager can delay the interruption until finished. This chapter presents the design and implementation of several classifiers based on implicit metrics necessary to build such disruption manager.

7.2 Text as implicit metrics

The Internet has given raise to several communication channels, making text the primary medium of representing and transmitting information. People work and communicate through email, instant messages, websites, and shared documents. Some progress has been made in textual analysis and it is now possible to extract useful information from textual information.

Large databases of commonsense knowledge provide semantic knowledge that has the potential to make sense of textual information (Singh et al 2002). Tools, such as ConceptNet support contextual commonsense reasoning and several practical textual-reasoning tasks over real-world texts (Liu 2004). Natural-language-processing and ConcepNet's commonsense knowledgebase were used to generate context-based inferences to describe text documents typically used on desktop computers. Spread activation of the commonsense semantic network provided conceptual relationships between documents and instant messages.

The work centered in the development of a program that analyzes a person's goals using context information, and ultimately determines if the content of monitored interruptions are relevant to their work. With this information, it would be possible to optimize effectiveness of the interface to support the intended goal of the user.

7.2.1 Concept topics extraction

The first stage consisted of analyzing emails, web pages and instant messages used on the experimental customer service scenario described in Chapter 6. Running each email and instant messages through ConceptNet provided a range of topics and common concepts (see Appendix E.1 for a list of emails used). Using this information, the commonsense database was customized to encompass a wider range of relevant topics pertinent to the emails and messages.

Each email, web page, and IM was automatically scanned and reduced to a list of words for faster processing. The list of significant words includes foreign words, adjectives, nouns (singular, plural) and verbs (past, present, 3rd person). These words represent the entire document. Contextual spread activation extracted a list of relevant topics/concepts with their corresponding scores. This list represents all relevant topics with percent relevance higher than 10%. In some instances, this list was further sifted to show only topics that are common amongst two or more documents. For example, this email was reduced as follows.

7.2.1.A Email (Email-quality1.txt):

Hello,

We will be running an exercise camp for teenagers this summer and would like to evaluate some sports products. If satisfied by the products quality and pricing, we will place a bigger order to equip the entire camp. All items are available on The Sports Authority Direct Catalog. Budget: $3,100.

Step System.
Men's bicycle.
Skate helmet.
Body Ball.
Waterproof cargo.

Connor Yang
Camp director

7.2.1.A.1 Document representation (reduced list of words):

running	*camp*	*summer*	*sports*
exercise	*teenagers*	*evaluate*	*products*

satisfied	*equip*	*authority*	*bicycle*
products	*entire*	*direct*	*skate*
quality	*camp*	*catalog*	*helmet*
pricing	*items*	*budget*	*body*
place	*are*	*step*	*ball*
bigger	*available*	*system*	*waterproof*
order	*sports*	*men*	*cargo*

7.2.1.A.2 Topic list representation

Document file: email-quality1m.txt

[('run', 0.032729278908523084), ('exercise', 0.026746826481981654), ('person', 0.025344280301956391), ('walk', 0.0076966726144291472), ('go jogging', 0.0064788304151501 76), ('car', 0.0054939356144777657), ('animal', 0.00577935672170851), ('eat', 0.0047111048228213733), ('water', 0.0053536781906696456), ('injury', 0.0044855358793343889)……]

7.2.2 Document Discrimination

Context-based inferences are effective in generating topic lists describing text documents. A list of top topics was assembled and used to identify communalities between documents. Documents, emails or instant messages were classified according to the topics used to represent them. Figure 24 shows emails classified as being relevant to the topic “furniture”. Instant messages are also classified according to their topic words. This classification is later used to correlate interrupting instant messages to the concepts surrounding active documents.

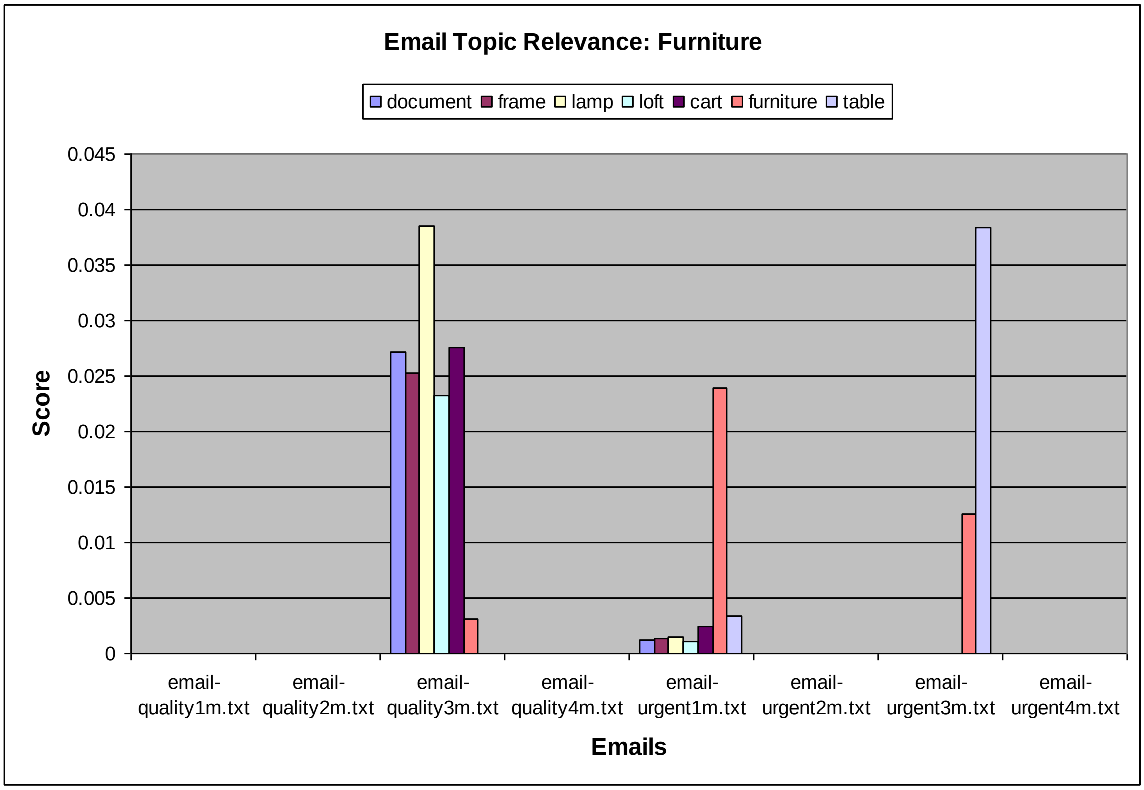


Figure 24 Email topic relevance classification. Five out of eight emails do not fall within the furniture category. This figure shows a limited number of topics to ease readability

On a more specific example, the email (Email-quality1.txt) previously used and the email (Email-quality3.txt – See next email) significantly produce distinct topic words. The topic words that represent each email are quite accurate: budget, catalog, run, ball, exercise, store, furniture and office. The topic words 'budget', 'product' and 'catalog' are equally represented in both emails, indicating that they share similarities. However, each email is also represented by their unique topics; Figure 25 illustrates these differences graphically . Using this data, assertion of the concepts surrounding the user's goals becomes possible.

7.2.2.A Email (Email-quality3.txt):

Hi,

We are an interior design firm expanding to new locations and are looking for a new provider to handle large Crate and Barrel purchases. Please provide competitive quotes for the following products available at the Create and Barrel Catalog. We will inform you of our decision based on your quote. Budget: $1,450.

10 Document Frame.5 Kyoto Lamp.

9 Loft Three-Shelf Cart.

Tiffany Graff

Design Acquisitions representative

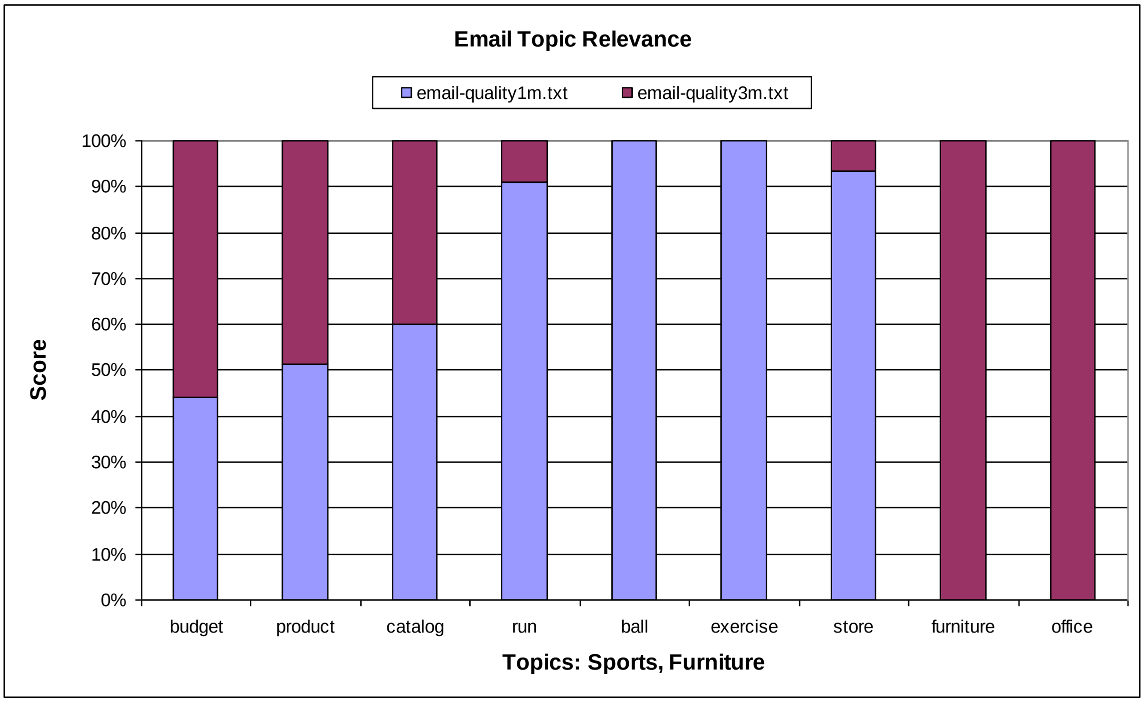


Figure 25 Email topic comparison for two emails. The first email (email-quality1.txt) evidently relates to sports, whereas the second email (email-quality3.txt) relates to furniture.

7.2.3 Relevance Score

The last stage of the work dealt with determining a score, based on the relevance between a text document, which for testing purposes was an email or a Google search result page, and incoming interruptions, such as an instant message. Two methods were used to calculate a score, the first of which relied simply on common words between the two pieces of text. A mathematical formula was used to calculate a relevancy score between 1 and 100. The formula included the number of words in text1 (email) that could be found in text2 (IM), total # of words in text1 (email), number of words in text2 (IM) that can be found in text1 (email) and total number of words in text 2 (IM).

The common word comparison of the two documents showed reliable consistency. While Instant Messages relevant to the emails generated values of around 10 percent, irrelevant messages generated values between 0 and 3 percent. Relevant Instant Messages compared to the Google search page results generated larger relevancy percentage values, while irrelevant messages generated values of 0 to 1 percent.

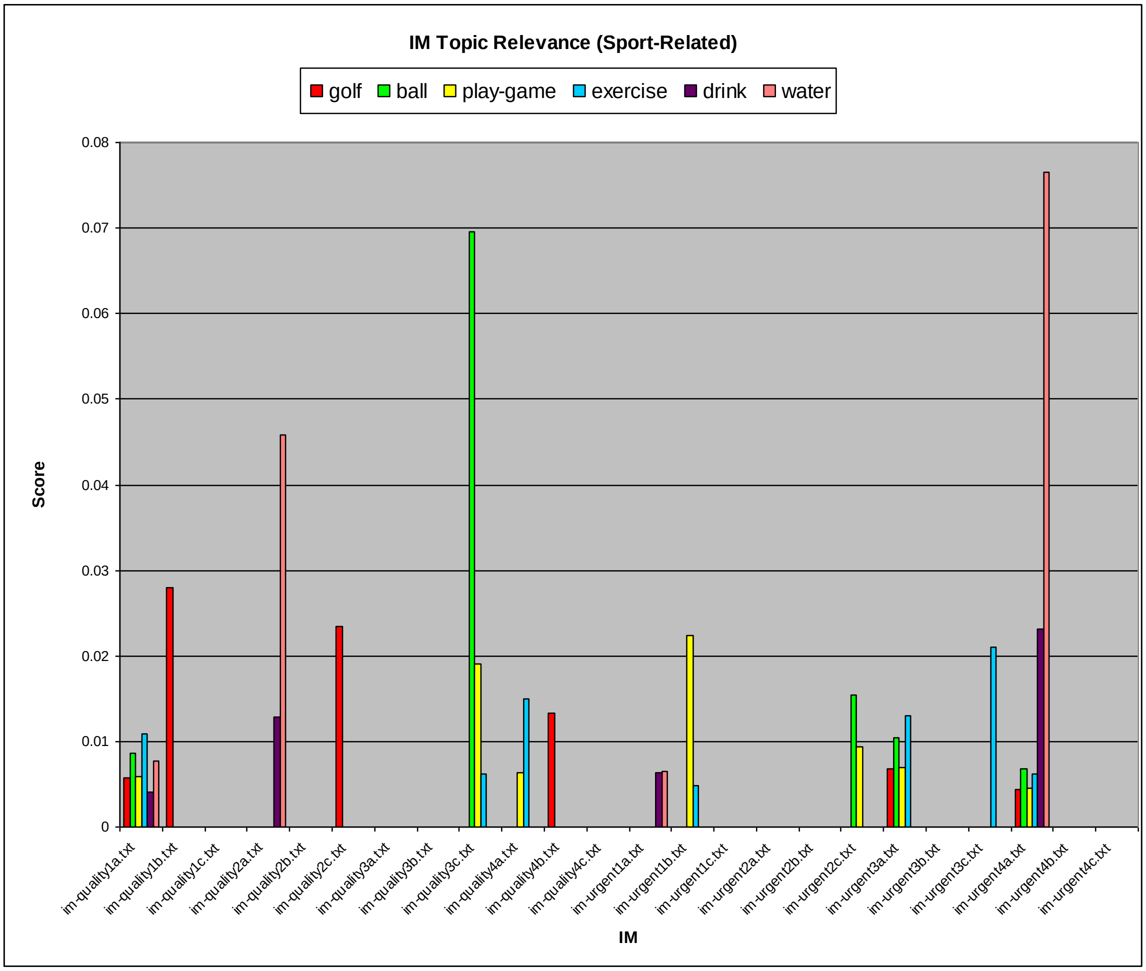


Figure 26 Instant Messages classified as Sports-related.

The second method of determining a weighted score relies on generating a list of possible topics with a percent relevance for text chunks using the topic-gist function in ConceptNet. Common themes and their percent relevance values are collected into a list. Entire instant messages are also processed in the same way. Finally, the percent relevance values between the common topics amongst both the text document and the message determines the final score. In most cases, the values of relevant instant messages presented significantly higher scores that those of irrelevant instant messages—higher discriminating power than simple common word comparison.

7.3 Mouse tracks as Implicit Metrics

Previous research and our investigations suggest that the mouse cursor can provide information about where the user focus of attention is. Salvucei also suggests that by predicting gaze position, we might be able to infer user's intent based on the mouse tracking data (Salvucei 1999). Furthermore, this information can also be used to classify user navigation behavior into several categories, such as scrolling, reading, thinking, or interacting with menus.

While glancing might imply attention, it is part of autonomic responses. Mouse movement on the other hand is the result of a complex conscious decision related to the goal of action. As such, it implies commitment to do something. This generally shows that conscious acts are especially good implicit metrics for computer interfaces, especially for recognizing the all important context switch that can interrupt or reestablish flow. This indicates that mouse movement could be an indicator of user activity and disruption. However, we must consider the possibility that the cursor control behavior is so practiced that it has moved down the cognitive hierarchy to a level at which people unconsciously make movements and actions with their mouse.

We have focused on mouse monitoring because it is the main interaction medium across several desktop platforms and can reveal useful information without the addition of new sensors for monitoring user activity. First, we explored mouse movement as indicators of user activity within web pages. Mouse tracks proved to be a valuable tool for web page design web usability testing. Later, we explored the idea that mouse usage provided enough information for people to understand the user work flow and decide on appropriate moments for interrupting. Having proved the concept of using mouse metric for understanding disruption, we evaluated mouse metrics even further on their ability to predict user interest and provide data regarding user activity, which in turn will be used to control disruption.

7.3.1 Application domain

We decided to focus on web navigation as according to Nielsen/NetRatings, 68.6% of American adults use the internet (Nielsen 2006); representing approximately 147 million people as of March/2006. Information search and retrieval is one of the main activities people do online as 91% of internet users report using a search engine to find information. Internet is becoming ubiquitous and length of time people spends online is increasing. As people spend more time online browsing for information, systems that allow this activity to go undisturbed are needed. Analyzing server logs can provide quantitative data on the success or failure of a website (Scholtz 1999). However, logs do not provide vital detailed information about what users are doing at a more granular level than the page-view level.

Server logs do not show where users struggle while browsing a website or what activities visitors perform on a website (Hochheiser 1999).

Website visitors often change their surfing habits overtime. As visitors gain familiarity with a site, they approach it differently. Surfing experts navigate in entirely different way from novices (using shortcuts, skimming through websites, searching for matching words, etc). A study examining, mouse movements on web pages found that 35% of people moved their mouse cursor while reading a webpage (Mueller 2001). Some web surfers move their mouse cursor according to their focus of attention, while others move their cursor to a blank or scrolling area while reading a webpage. This suggests that the mouse cursor is an indicative to where the user focus of attention is. For instance, a relatively low traffic site with 1000 visits per day could potentially provide 250 data samples for analyzing how people navigate through the site. Furthermore, the analysis could be complemented by online surveys that can reach a large number of people, adding statistical power to the results.

7.3.2 Related Work

Studies have explored mouse movement as an alternative method for cognitive studies in language comprehension. For instance, mouse movement trajectories could indicate how information is interpreted. Slow and arched trajectories as users move their mouse would indicate an ambiguous state of mind (Spivey 200). Eye-tracking uses several indicators of ocular behaviors, including fixations, saccades, pupil dilation, and scan paths (Rayner 1998). Studies evaluating these metrics have shown that there is a correlation between users' attention, their eye movements, and their cursor movements while browsing a webpage. These studies suggest that there is a strong relationship between gaze position and cursor position (Chen 2001).

The Enhanced Restricted Focus Viewer (ERFV) was developed as a software-based alternative to expensive eye-tracking systems. Its mouse-based approach collects the path of users' visual attention as they browse a website without the need of specialized hardware. The ERFV has been used to analyze web browsing patterns on several sites (Tarasewich 2004). The ERFV requires individual images to be generated for each web page. This manual intensive requirement makes it unsuitable for large websites. A major drawback is that users might not browse the web as they would normally do due to the restricted field of vision, which is limited to a small focus window.

Studies linking web usage to eye movement behavior indicate that web page viewing behavior is determined by gender, order of web pages being viewed, and the interaction between site types and the order of the pages being viewed (Pointer 2000, Josephson 2002, Goldberg 2002, Hseih-Yee 2000, Pan

2004). Gender differences in perceptual processing indicate that females often engage in comprehensive processing of all available information, while males tend to focus their attention on fewer areas—males often exhibit longer mean fixation time than females (Jones 1998, Meyers-Levy 1991).

7.3.3 Mouse Tracking Tool

Mtrack is a visualization tool that displays the mouse path followed by website visitors. The path is augmented with arrows indicating directionality. In addition, the webpage entry point is clearly marked with a different color. Visualization includes a shaded area around the cursor that increases its size to represent time spent over a point. Shade intensity also increases to represent mouse hesitation over continuous zones. The tool allows web pages to be divided in different "to-be-monitored" areas (see Figure 27). The areas are highlighted based on the mouse activity in/or around them. The degree of activity to each area is reflected by shade intensity; similar to a heat map.

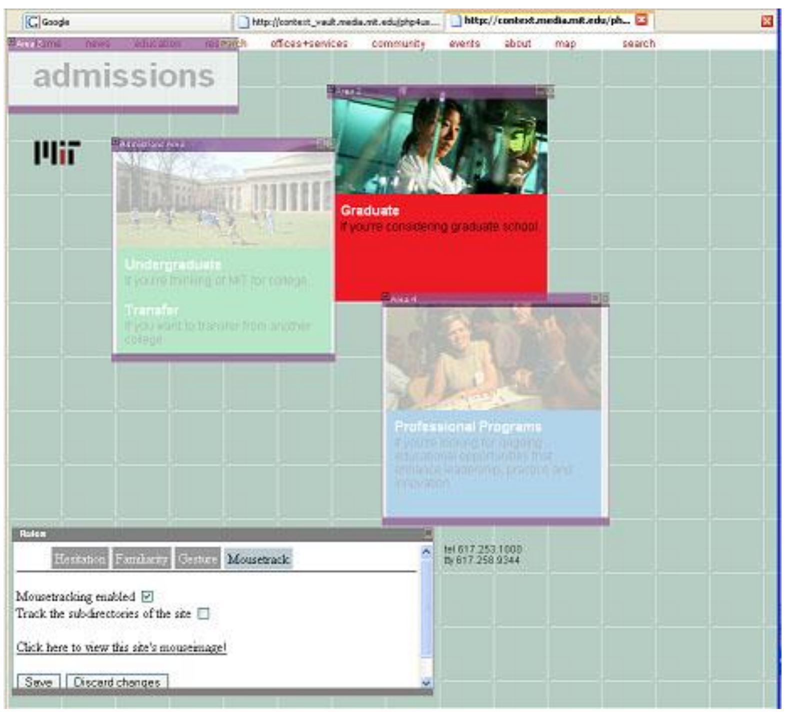


Figure 27. Pre-defined areas on a web page to be tracked. They can be minimized for visibility.

The visualization tool aggregates information, but it also allows information up to the session level to be evaluated. A floating configuration window on the analyzed webpage provides multiple visualization options. A unified interface consisting of a pair of sliders allows the administrator to control the amount of information to be displayed by clusters representing time frame, IP address range, and sessions.

7.3.3.A Usability implications

Mtrack offers advantages when compared to traditional usability tools. First, it can be mass deployed, allowing for large datasets. Second, it is able to reach typical users and first time visitors in their natural environment. Third, it can continuously test live sites, offering insight information as new website sections are deployed. Fourth and most importantly, it is transparent to the users, so no experimenter bias or novelty effects are introduced, allowing users to navigate, as they would normally do. Existing web-tracking usability systems, such as WebVIP, WET, and WebQuilt focus primarily on logging mouse click events/interactions (Etgen 1999, Hong 2001, Scholtz 1998). In contrast, our system focuses on mouse browsing paths within a webpage.

The nature of the tool has the disadvantage that not all data offers information about specific user interactions with a webpage; as not all users use the mouse as a reading aid. However, long resting positions followed by rapid linear movement could be an indicative of visual focus not determined by the cursor. The tool allows administrator to identify this type of situations by visual inspection. Cases where not enough data points are available are automatically labeled for revision and excluded from the visualization. Another limitation is that the system/tool might interfere with web pages that do not conform to W3C standards. It may also interfere on pages that rely heavily on scripting tools for their appearance and/or navigation.

7.3.3.B Dynamic Personalization

Mtrack goes beyond providing usability information; it has the capability to dynamically augment websites based on detected browsing behaviors Web pages can be personalized as response to users' web browsing mouse activity. Mtrack offers a basic but powerful set of configurable parameters to create rules. Website administrators can make use of these rules to implement different marketing schemas/responses based on identified users' interest and administrators' individual goals. Some examples are promoting material that is rarely seen, personalizing options based on familiarity, making frequently used links readily available, making information that the user might be interest in more prominent. It is up to the administrator to generate appropriate rules for their site and decide what type of rules will be available.

A rule consists of two components: a precondition and a recommendation. The precondition is based on events related to mouse activity, such as the user hovering over a link or an area previously defined for mouse tracking. A recommendation is a suggested link and/or image that appear on either the current page or a subsequent page. Suggestions are embedded on the webpage as a frequently used links section, similarly to the "Recently Viewed Items" window on amazon.com

The systems implements two specific types of rules: hesitation and familiarity rules. A hesitation rule is based on how much time the user spends over a link or an area relative to others without clicking on them. (flying by an area or link is not considered hesitating). Mueller suggests that "hesitation on links could potentially provide information about users' other interests on a webpage" (Mueller 2001). A familiarity rule is based on how much time the user spends in an area before clicking on a link. The administrator may create a rule that if the user is deemed unfamiliar with the site, a suggested link would let the user know the existence of a tutorial page in the web site.

7.3.3.C Implementation

Mtrack was implemented as a web application that can be accessed through a web browser. A PHP proxy fetches existing web pages and modifies them by inserting JavaScript code. Returned enhanced versions of these web pages implement an administration interface and add mouse tracking capabilities when browsing a webpage. The end-user can browse enhanced web pages as they normally would, while Mtrack transparently records mouse activity within the web page.

The mouse movement data collects samples whenever a user moves out of a 50 pixel circle radius and logs the time spent at each position/coordinate. This distance sampling approach filters out very fine movement to reduce data size and remove unnecessary information. When the user clicks on any link, the full coordinate set (2.5k on average) is sent as a request to the proxy server, which in turn stores the data in a database and redirects the user to the desired link. The tool does not require changes to existing websites, and no expert configuration is needed. All that is necessary is a web browser, an internet connection and the URL for the desired page to be tracked.

7.3.3.C.1 Administration Interface

The administrator interface is designed for administrators to configure mouse movement tracking parameters, visualize mouse tracking data, specify links or regions of interest, and set up rules on existing web pages. Rules specify what recommendations are displayed based upon a condition on mouse movement activity. The administrator interface retrieves existing web pages augmented with a floating configuration window.

7.3.3.C.2 End-User Interface

The end-user interface is intended for users to browse through web pages enhanced with mouse tracking capability and link recommendations. The tool does not require changes to existing websites. A proxy fetches existing web pages, modifies them by inserting mouse tracking code and recommendations and finally returns an enhanced version of these web pages. The sole difference from the original web pages are embedded recommendations generated based on the rules previously configured by the administrator. The recommendations for rules based on mouse activity on the previous page are static, while the recommendations based on mouse activity on the current page are updated as the user mouse navigates on the page.

7.3.3.D Design case

Practitioners used the admissions webpage of a university as a test base. This website was selected because is simple enough so that enhancements can be easily understood and implemented. The site is divided into three main topics: undergraduate, graduate, and professional programs. Additionally, the site has a menu linking to other university sites. The professional programs section of the webpage is rarely visited. The menu causes users to go back and forth web pages since many of the links are external. The undergraduate link is the most popular. Figure 28 shows several sessions from the same user over a period of two weeks. During the firsts sessions the cursor indicates that the visitor explored around the website before clicking on a link. In contrast, the last session shows a rectilinear path from the point of entry directly towards the undergraduate link. This indicates familiarity with the site.

Based on these observations, the site was configured as follows: An area was defined for each section of the admissions page. Each menu link was associated with a thumbnail image of the destination page. The image in turn was associated to a suggested link to the menu link itself.

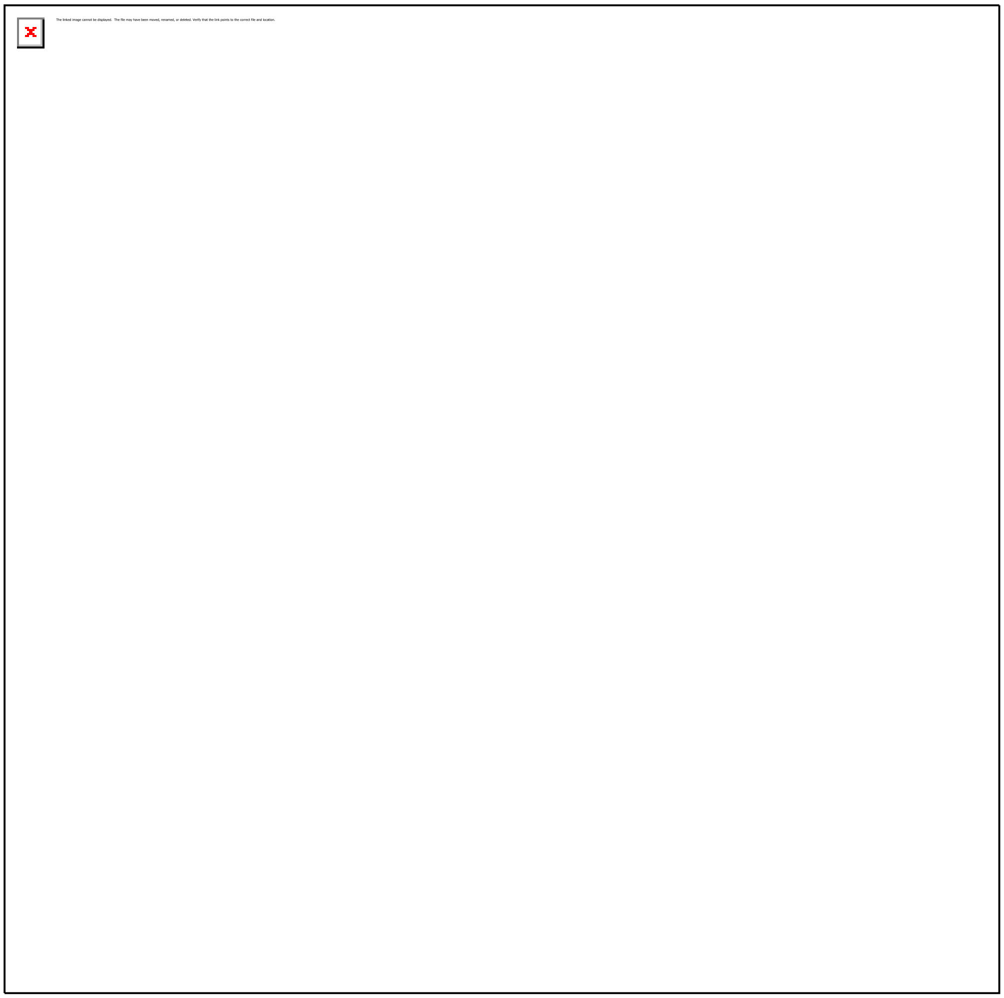


Figure 28 First and last three session of a visitor to a website. The last session shows a rectilinear path from the point of entry towards the desired link.

Three hesitation rules implement a navigation aid. The rules specify that the top three most hesitated links on the menu should trigger a suggested link-image. Thus, visitors will see a preview image of the site that may follow. Two hesitation rules were defined in order to promote the professional programs section. These rules trigger a suggestion for the least hesitated link between the undergraduate/graduate areas and the professional programs area. If the user moves the mouse around areas on the graduate and undergraduate programs and the professional program links are ignored, a suggestion to the ignored links will appear on the next page. And finally, a familiarity rule for the undergraduate area detects if visitors have been to the site before and presents them with a suggested link to the academic programs webpage. Theses rule might not be correct all the time, but when right, the enhancements could help visitors find their way through the site.

7.3.3.E Initial findings

The system was introduced at the Asian Reality 2005 workshop to a group of architect practitioners and students that had little or no experience in HCI and interaction design. The tool made it possible for the students to perform web page explorations on 7 different websites. Their analyses included music reviews, dating, museum, travel, and design competition websites. Websites varied from highly graphical to text only sites. A total of 105 subjects were asked to navigate these websites as they would normally do. They were later instructed to describe what they did on the site. Finally they were presented with their own mouse trajectories and further described what they meant according to the previously described behavior for the site.

The visualization approach of the tool, made the issues and possibilities for improving websites obvious to the workshop participants. The students very quickly went from running subjects through an experiment to being captivated with how web page design changes the way people think. They proposed and prototyped redesigns reorganizing information where it could be easily found, and simpler to navigate.

The system provided mouse activity that was useful in understanding how users viewed a webpage and aided in identifying potential problems with a webpage. For instance, when evaluating a design website, participants found that activity concentrated on three main areas: website logo, website description and selection menu (see Figure 29). Even though this is a simple webpage, there was some confusion about the site's interactivity. Most people attempted to interact with the website logo on the webpage, but it did not have an interactivity associated with it. Interestingly, the short website description was effective in getting people to read about it, being one of the most viewed areas. And finally, they discovered that people spent most of their time interacting with menus and selecting art designers to review. This was expected since menus are the only interactive part offered by the site.

Figure 29 Three areas in a webpage showing the most activity from first-time visitors to a design museum website.

Mouse movement data across websites can be classified into several behavior categories: scrolling, reading, pause-think-read-go, interacting with menus (graphical and text only), and random. Scrolling is easily identified by mouse trails up and down the scroll bars. Reading is indicated by smooth mouse trails movements horizontally and vertically across web page paragraphs (see Figure 30). Not all people move their mouse while reading, instead they park their cursor to either read or think about what to do next. This behavior is characterized by a long pause next to some text or a blank space and fast and direct movement towards a link, usually terminating with a click. Interacting with menus is characterized by hesitation among the different menu options. Finally, random movement refers to people playing around with the mouse cursor in no predictable way and with no underlying motivation.

Figure 30 Mouse trajectory indicating reading behavior and hesitation in the menu area. Medium, closely spaced circles indicate smooth movement through the paragraph

7.4 Mouse Tracking for Mediating Disruption

This section describes exploratory work into the relationship between mouse, user attention and interruptions. Several studies and demonstrations investigate the feasibility of using mouse tracking for different purposes, other than web usability. A first user study investigates interruptions while performing online tasks. A second one shows a user study designed to evaluate and compare mouse-tracking with traditional eye-tracking studies. This study produced a simple activity demonstration that illustrates several activity recognition concepts. A third study evaluates opportunities for detecting behavior and predicting user interest by evaluating mouse activity. This section also presents a system designed to perform classification tasks based solely on mouse tracks. Finally, the section discusses some applications and implications within the study domain.

7.4.1 Mouse activity as perceived by users.

This study tried to identify clues that people use when interrupting another person. Typically, people use cues for determining appropriate moments for interruption. If somebody just finished taking on the telephone; that action might be a good indication to knock at their door. The study investigated if humans are capable of inferring interruption times with an experiment that presented a series of screenshots from people navigating websites.

Informal explorations showed that people were able to identify or guess what a computer user might be trying to accomplish based solely on screen captures including the user mouse activity with no further context. Furthermore, people used their own intuition regarding the most appropriate times to interrupt/distract the user whilst performing a task. Mouse movement trajectories could indicate how information is interpreted while browsing a webpage. For example, slow and arched trajectories as users move their mouse would indicate an ambiguous state of mind.

On an online survey, we asked subjects to identify the times they would choose to interrupt someone when performing a task on a computer. Subjects were presented with several videos from people performing a task on several websites and asked to select three possible interruption times based on mouse activity. The videos presented included screen captures of people booking a flight, and finding directions. These three activities were performed on three different sites that offered similar functionality. See Table 2.

Booking a flight	Driving directions
Expedia.com	Maps.google.com
Cheaptickets.com	Maps.yahoo.com
Usairways.com	Mapquest.com

Table 2 Booking a flight and finding directions tasks were performed on the following sites.

The tasks can be described as a series of intermediate subtasks: Decide-Commit, Data-Entry, Wait, and Evaluate. The 1st subtask includes a decision stage, where people decide what to do before taking an action, i.e. making up your mind about which search engine to use to find an airline website. The 2nd subtask involves entering and submitting data. The 3rd subtask involves waiting for information to be presented, such as waiting for a webpage load. In addition, the 4th subtask involves evaluating information, such as evaluating the search results webpage. The subtasks are performed as a loop through the duration of the task. The task video shows people executing the same subtasks with different parameters. Appendix A.1 illustrates the subtasks cycling nature.

7.4.1.A Protocol

Subjects were asked to identify the top three interruption times and to explain the rationale behind their decision. Subjects had the ability to replay the videos as many times as necessary and to change their interruption times. The survey interface provided the required flexibility and ease of use for selecting and updating multiple interrupting times. This survey provided insight about the type of approaches people might use when dealing with interruptions. Subjects are computer users themselves, and their expectations regarding interruptions might be related to the way they approach/interrupt someone while working.

7.4.1.B Results

Participants' approaches fell within two categories: high level and micro level approaches. High level approaches had to deal with users accomplishing their goals, why are users doing something and asking questions about whether or not the user had accomplished his goal, i.e., "This is an okay time to interrupt because the person has found their location already so they don't need to concentrate on getting the right address". 82% of people identified the end of a task as a good moment to interrupt someone, where as, the start and middle of the task were selected only 6% and 12% of the time.

Micro level approaches are characterized by issues related to the user interface. The user interface signals appropriate moments for interruption (typing in text boxes, pressing a button, and so on). For example, "User has finished typing for the moment. User is waiting for the webpage to load and has a free moment. The user can be interrupted at this time and be able to quickly resume looking at the results when they load".

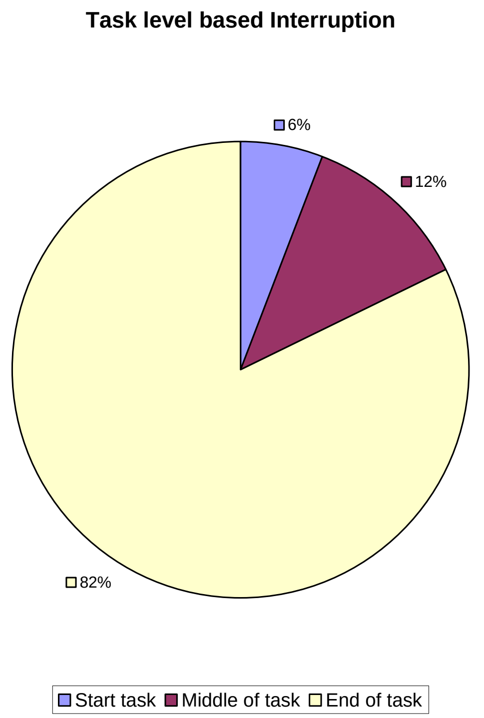


Figure 31 Subjects approaches to interrupting someone while navigating the web

7.5 Mouse Tracking Feasibility Study

New research exploring the relationship between eye movements and mouse movement suggests that there is a strong relationship between gaze position and cursor position (Chen 2001). However, research also indicates that the relationship between cursor and gaze position (Byrne 1999) will vary sometimes. There was a need to evaluate eye movement with mouse cursor movement at a more granular level rather than dividing a webpage into areas of interest (an approach commonly used to aggregate data). There is extensive research on eye-tracking evaluating web usage (Goldberg 2002). Some work of particular interest is the research focused on tracking user behavior while examining web search results. Their results show that people pay attention to the 1st - 3rd search results and more likely to click on them. Attention for the remaining results decreases drastically on the 6th and 7th results. The type of granularity offered by users examining web search results is appropriate for evaluating the relationship between mouse movement and eye gaze. As web search becomes widely used as the main access point for information retrieval, it is important to understand how people interact with search results.

This study illustrates the relative value and tradeoffs between mouse monitoring by replicating and extending results from eye-tracking research. The study focuses on web search models and web page navigation and validates the use of mouse tracking as a valid tool for understanding users' attention and browsing behaviors while evaluating web search results. It also proves the advantages of using an unobtrusive tracking system.

The study used the mouse tracking tool described previously in Chapter 7 with no major modifications. The experimental tool monitored and logged subjects' mouse movement activity as they interact with web sites. The tool also allowed for visual inspection of subjects' interactions within each webpage. Visual inspection proved extremely useful in evaluating the types of user interactions with search results.

7.5.1 Protocol

Participants were presented with the Google home page and performed several web searches until satisfied with the results. As participants complete their task, their mouse movements were recorded by our tracking tool. Participants were presented with three different types of search tasks:

- A free search, where participants can search for anything they want.
- A prompted search, where participants can search based on pre-defined scenario, using search queries that seem appropriate.
- A simulated search, where participants can evaluate a simulated set of results from the same query (also based on a pre-defined scenario).

Search scenarios include two categories: one scenario related to finding online information about people, and another scenario related to finding information about commercial products. Searches related to finding people did not include sponsored links, whereas searches about commercial products included sponsored results.

The experimental software filtered search results in order to achieve a fully-controlled experiment and to maintain consistency within searches. The software filtered out OneBox results and Google Sponsored links from the search results; see section H in Figure 32. OneBox results are typically included at the top of the search results and include news, stock quotes, weather and local websites related to the search. Sponsored links are paid by advertisers to match terms related to the search.

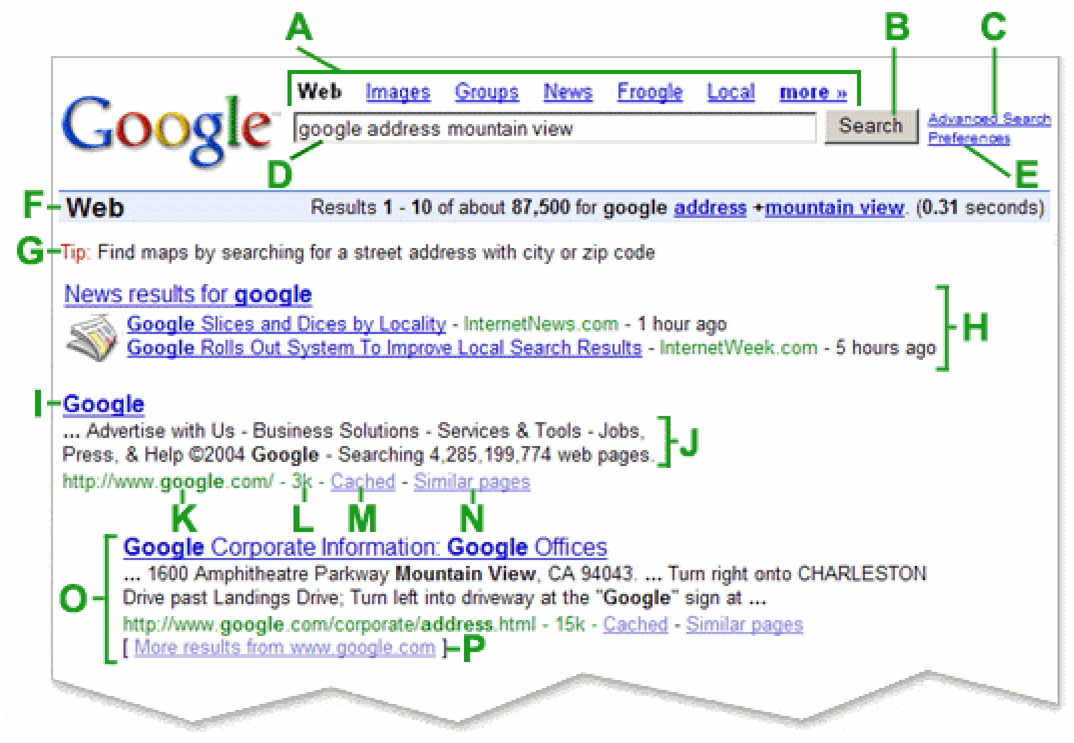


Figure 32 Google Home Page Description (http://google.com). Areas in the results page are indicated by letters A trough P. See Appendix C.2 for a full description of each section

The order of presentation for each of the search tasks was counterbalanced to avoid any presentation effect. Participants were randomly assigned to one of three counterbalanced conditions; see Table 3. Each participant performed 6 searches total.

	CONDITION		
	1st	2nd	3rd
1	Free Search	Prompted Search	Simulated Search
2	Free Search	Simulated Search	Prompted Search
3	Prompted Search	Free Search	Prompted Search
4	Prompted Search	Free Search	Simulated Search
5	Prompted Search	Simulated Search	Free Search
6	Simulated Search	Prompted Search	Free Search

Table 3 Counterbalanced search scenarios.

7.5.1.A Search Tasks Examples

Several scenarios described in general terms the information that subjects should search for. These scenarios included feasible descriptions that people could easily relate to and quickly understand the objective of the search. See Appendix C.1 for a comprehensive list of examples.

"You are interested in contacting an old professor to ask for a recommendation letter. Eavan Boland is a poet with whom you worked for some time, but you lost track of her. You should find out what she is up to now and get her email address."

"You are interested in buying tires for your Toyota corolla and want to find information about what tires you should buy. Assume that you live in Miami and want to find a local store that carries the tires you selected, as well as store their store hours."

7.5.1.B Procedure

The experimenter presented participants with a questionnaire designed to assess computer experience level and familiarity with the web browsers. Before each session, the web browser's history and cache information were restored as not to bias participants with "marked" hyperlinks or queries entered by previous participants and to maintain consistency through the experimental conditions.

The experiment was conducted on an isolated laboratory room under the same conditions for all participants. The experimental setup included laptop sitting on a standard office desk with adequate support for the participants' arms and a standard office chair. The computer had a secondary monitor, keyboard, and mouse, which allowed the experimenter to monitor the participants' mouse activity and was useful in setting up the computer settings promptly. The computer used the same resolution for all conditions (1024x768), which is common on today's computing environments and it doesn't require excessive scrolling due to limited screen size.

7.5.1.C Participants

30 subjects participated in the experiment, 50% male and 50% female with ages ranging between 17-30 years old. Subjects' familiarity with web browsers varied from intermediate to expert.

7.5.1.D Metrics

Thorough the duration of the experiment, the experimenter monitored the user behavior and manually labeled the participants' mouse activity over time in approximately 10 seconds increments. The experimenter also visually identified if the participant's mouse moves on each webpage might be adequate for mouse tracking.

Data related the web page search results and other metrics included type of task, Subject ID, Query used search results page access times, general observations and activity, reading, quick exploration, scrolling, cursor over links, cursor static (limited cursor usage), and cursor moving erratically. The experimenter also labeled each session based on subjects' search behaviors as quick search, linear search, golden triangle search, deliberate scan.

7.5.2 Results

86 percent of participants used their mouse in an identifiable pattern, such as pausing and moving straight towards a link, moving their cursor as they were reading the search results, or scrolling slowly while reading. On some situations, expert subjects relied heavily on keyboard shortcuts to navigate the search results webpage and their searches were excluded from the dataset. The resulting data was filtered based on the amount of information that was logged for the participant and based on the correlation between the experimenter notes and a second observer.

The free search scenario did not provide enough context information for subjects to be able to decide on a search query. In fact, most people had trouble coming up with a search term in the free search scenario. Another drawback from this scenario is that some people knew specifically what they were looking for, while others entered random search terms. Data from the free search scenario was discarded due to errors in variance

The experiments were focused on tracking user behavior while examining web search results. The results rely heavily on imagery or presentation. These images reflect examples of some of the data collected in the experiments, aiding the interpretation of the results. Results are divided in three sections: Entry Point, Golden triangle, and Scrolling.

7.5.2.A Eye-tracking similarities: Entry point

The results show a well defined area where users start their journey on the results webpage. The entry point for the mouse cursor falls within a well defined area of approximately 300x 100 pixels. This area is located about 180 pixels below the top of the page, next to the first organic search result. It is important to notice that there are some variations on the elements that fall within or close to the entry area due to extra links introduced on the results webpage (depending on the number of sponsored links or suggestions included).

By overlaying the Google homepage with the search results webpage, we can see that the entry point lines up with the Google search button that used to be on the previous page; see Figure 33. The entry point is of special importance since it determines the first item that people would pay attention to. Although not surprising, eye-tracking studies have documented the initial fixation of the eye in the results webpage and have shown that this is the first place that people pay attention after the results page has been loaded; see Figure 34.

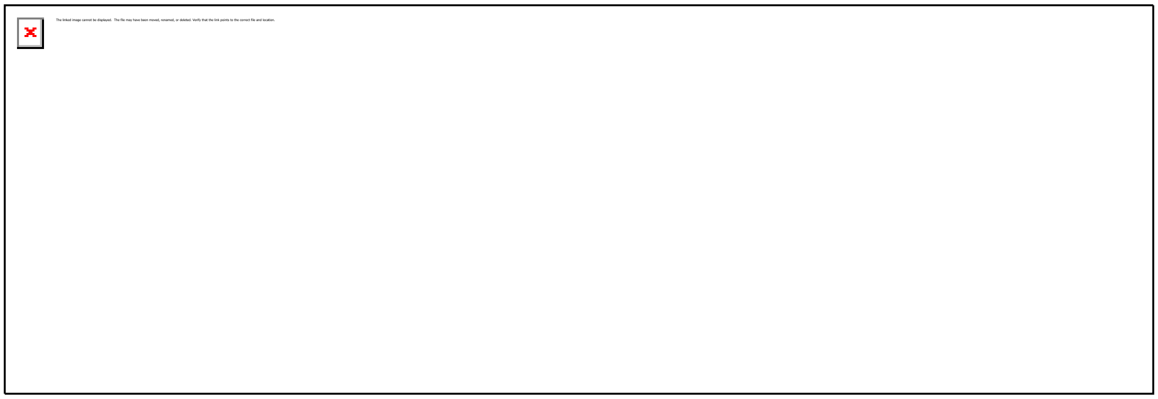


Figure 33 Image shows home search page superimposed over search results. Green dots show cursor entry point.

7.5.2.B Eye-tracking similarities: Golden Triangle

There is extensive research on eye-tracking evaluating web usage (Enquiro 2005). Some work of particular interest is the research focused on tracking user behavior while examining web search results. Their results show that people pay attention to the 1st - 3rd search results and are more likely to click on them. Attention for the remaining results decreases drastically on the 6th and 7th results.

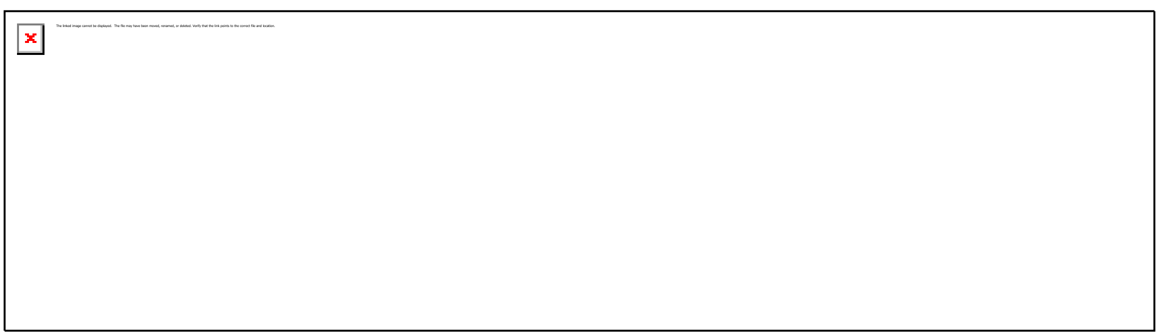


Figure 34 Google search results heat map. Hot Spot shows entry point of eye (Enquiro 2005)

7.5.2.C Eye-tracking similarities: Prompted Search Vs Simulated search

Eavan Boland Vs Robert Jacob

These searches provided the ideal circumstances to show any condition effects. Search results were limited to a few links and no extra sponsored links that might interfere with the users' decision process.

We confirmed the "there's something hidden" effect. Subjects on the simulated search condition explored the search results looking for a magic link. They were expecting to find an artificially placed link. This effect can be seen on the image. Total mouse spread activity within simulated searches show that people were more likely to look for a magic link within the results. People took longer to decide which link to click.

Visual examination showed a significant difference on subjects' decision time (number of seconds before activating a link). Subjects' behaviors were significantly different between prompted and simulated searches. 70% of subjects on the prompted condition activated a link within a few milliseconds, compared to 25% on the simulated condition. People trusted on their abilities to come up with optimal queries, which is also reflected on the tendency to select the first search results without reading other results.

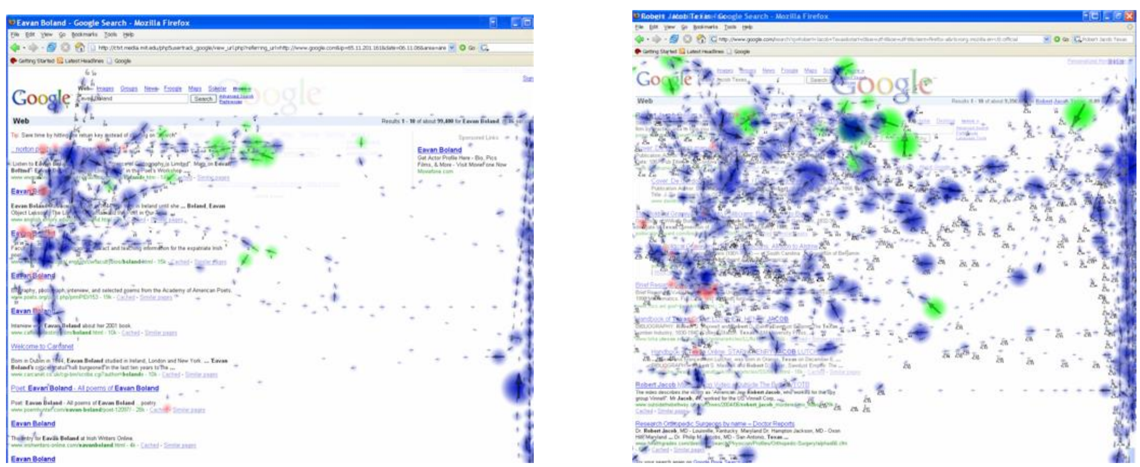


Figure 35 Differences in the simulated and prompted condition confirm the bias effect introduced by the simulated condition. These differences also indicate that people trusted their own search queries.

Toyota tires Vs. Digital Cameras

No differences in subjects' decision time were identified. About 50% of people activated a link within a few seconds for both conditions. This can be explained by the extensive number of results returned the queries on each search condition. Product searches yield sponsored links as well as commercial links. Polluting the screen with additional links and frames made the results harder to process.

Clicking behavior related to "Toyota tires" (prompted search) show that users were more likely to click within the 1-3rd search results, whereas the links activated for "Digital Cameras" (simulated search) varied with users activating up to the 6th search result.

7.5.3 Conclusion

The results validate the use of mouse tracking as a valid tool for understanding users' attention and browsing behaviors. The experiments replicated results from previous eye-tracking studies indicating that people pay attention to the 1st - 3rd search results and are more likely to click on them. Attention for the remaining results decreases drastically on the 6th and 7th results (Google triangle).

7.6 Implicit mouse metrics Proof of Concept Application

The previous study provided enough data to be able to develop an application that demonstrates the concept of a system that recognizes online activities in real-time. This section presents a simple classifier designed to detect scrolling activity from mouse tracks in real-time.

Traditionally web usage analyses are performed off-line and over massive data logs, resulting in difficulty to implement algorithms in real world scenarios (Hochheiser 1999). We have developed a scrolling application that offers a trade-off between model, feature, and computational complexity to achieve real-time performance. Figure Figure 36 shows data points for people classified as reading slowly while scrolling.

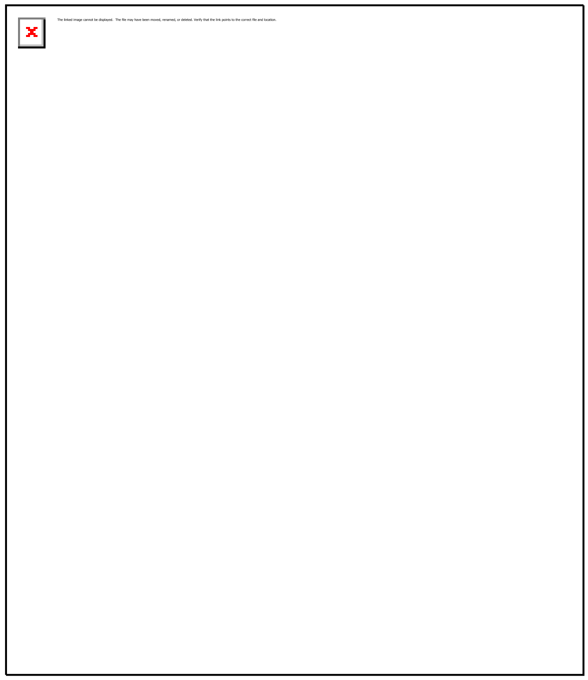


Figure 36 Data collected while performing searches on the Google webpage. Activity classified as scrolling (right).

The protocol was slightly modified and experimenters labeled entire sessions as mouse intensive or no mouse at all. Tracking and manually labeling subjects as previously done proved too difficult for the experimenters. Further labeling of the data was necessary to correctly identify user behaviors. The most suitable behavior for offline labeling was scrolling, since it presented agreement among several labelers.

In order to allow the activity recognition algorithm to produce classification outputs at semi-continuous intervals (time windows), a decision tree algorithm was used for this demonstration. Decision trees often generate understandable rules, which are easy to implement on the limited resources offered by web browsers. Decision trees represent activities in classes that are assumed to be mutually exclusive. They offer a binary "yes"/"no" determination for user activity, which is suitable for our demonstration purposes, however one major drawback is that they might over-fit the data.

7.6.1 Feature Extraction

Transformations from raw mouse activity data were used as features to predict several user activities. Features were selected based on their discriminating power and on their ability to compute them efficiently or inexpensively. A scrolling classifier was implemented as a J.45 tree. The classifier was implemented using a Java based pattern recognition tool kit (Witten 1999). Evaluation based on a 10 fold showed that the algorithm accuracy was 84%. A direct application from the preceding study is in search results evaluation. Figure Figure 37 illustrates how web search results can be evaluated based on users having trouble finding proper results. Search results relevancy can also be determined based on user's dwell time.

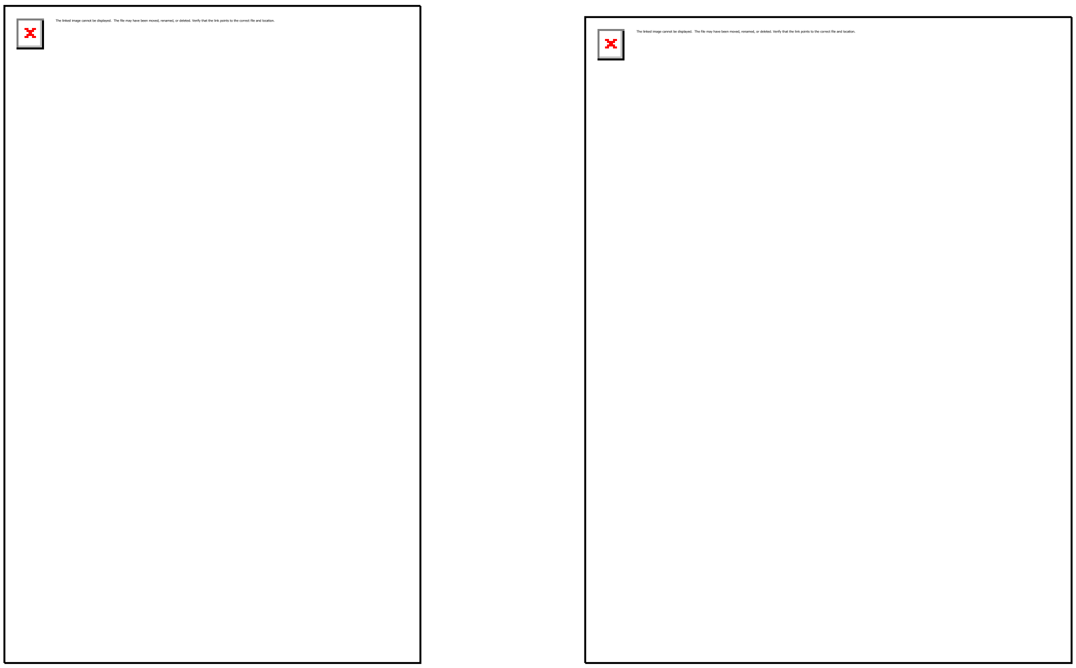


Figure 37 Differences in navigations of search results. Results not relevant to the search query (left). Results relevant to the search query generate distinctive mouse trajectories (right)

7.7 Online User Activity Classification from Implicit Metrics

As web search becomes widely used as the main access point for information retrieval, it is important to understand how people interact with the information provided on web pages and whether people are satisfied with the information they access. This experiment is designed with two goals in mind: first, evaluate the relationship between explicit ratings of user interest with implicit measures of user activity; second, gather training data for predicting user's interest and activities from mouse cursor implicit metrics. As previously demonstrated, implicit metrics for mouse tracking can indicate user's activity. The aim is identifying people's activities while browsing the web in order to inform a disruption manager. The manager would mediate based on user activity and user interest (on related concepts).

7.7.1 Mousetracking Labeling System

Modification to the tracking tool was necessary to include other metrics related to the webpage. These metrics include the elements in a webpage and time spent on each of these elements. A new experimental labeling tool was developed to automatically label subjects activities. The labeling tool allowed the experimenter to focus on observing the subject's behaviors. Whenever the experimenter identified a change in activity, he/she would label it through an easy-to-use point and click interface. The labeling tool would automatically link the logged data from each user with the experimenter's annotations.

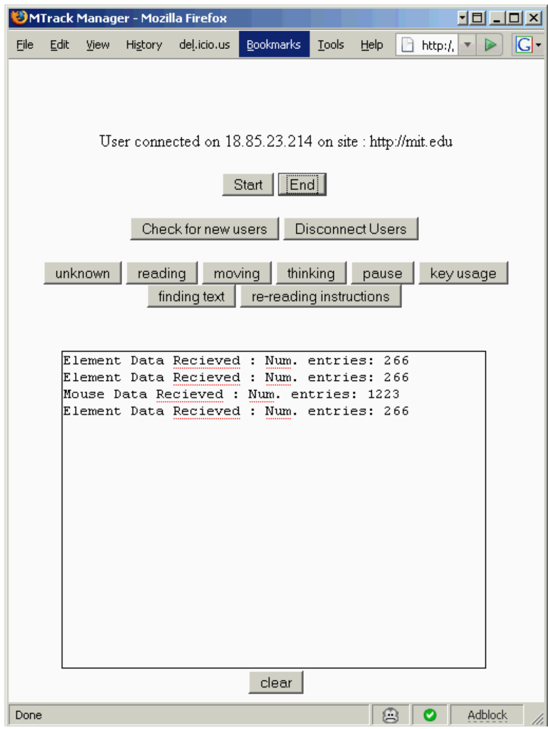


Figure 38 Experimenter's interface: labels are automatically synchronized to the user's data.

7.7.2 Pilot Study

The previous mouse tracking study validated the experimental procedure. However, a pilot study was required so that the experimenter could develop a consistent mental model of user activity and to test minor changes in the protocol. The pilot allowed the experimenter to standardize the labels used for each observed behavior; reaching a relative labeling consistency after 8 subjects and several practice trials. The pilot study also allowed the experimenter to familiarize him/herself with the labeling interface, and gain proficiency in labeling behaviors.

The previous experiment offered some insight into what type of user activity information could be relevant and the problems associated with collecting activity information. Having the experimenter focused on a single type of user activity (reading) assured dependable information.

A set of labeling guidelines guaranteed consistency while labeling user activity. These labeling guidelines include observations and updates from the pilot study; see Appendix C.4 for more information. The data and labels collected included low level details when possible, such as scrolling or using keyboard shortcuts (using arrows to navigate and Ctrl+F to search for information); although the task description instructed subjects not to use keys and read thoroughly. The data also included information about user activity:

- Reading without using the cursor as reading aid,
- reading while moving the mouse cursor as reading aid (following text laterally),
- slowly scrolling vertically (with or without use of the scrolling bar),
- pausing to thinking or collect thoughts,
- re-focusing attention (re-reading the paper instructions provided or getting distracted).

7.7.3 Experimental Design

The experiment evaluated people's navigation behaviors under to two randomly assigned conditions, a reading and a scanning condition. The reading condition presented subjects with several web pages and instructed them to read them thoroughly, whereas the scanning condition instructed subjects to scan them.

The experiment included five web page types; pages with link lists only, pages including short and medium summaries, pages with long articles, and pages including images (Links List, Short summary, Medium Summary, Articles, and Images). Links only pages required subjects to select a link from a long linear menu list and from a website site map organized by categories. Pages including short and medium summaries required subjects to select the most interesting articles based on short descriptions and summaries from an online blog and a magazine. Pages including images required subjects to examine several photographers' galleries and decide which one to hire and selecting a picture from an online picture gallery. Pages with long articles required subjects to read two online articles from a news website. The experiment included two web pages for each of the webpage categories and the order of presentation was counterbalanced through the experimental conditions.

7.7.3.A Protocol

An introductory task required subjects to navigate the MIT webpage using FireFox Mozilla web browser for a few minutes, until familiar with the web browser; if not already familiar with it. Subjects were instructed to examine several webpages and determine if they found them interesting. The experiment required subjects to visit several pre-defined and instructed to identify the top-three interesting articles, pictures, and links within those webpages. The instructions included short scenarios relevant to each of the web pages being navigated. The scenarios included a task to be performed on each page and varied depending on the experimental condition, see 0. The following is a sample scenario and task for an image website:

> *"A friend is moving into a new place and asked you buy an art piece for his apartment. He pre-selected some art pieces, but would like your opinion. Please browse an art webpage thoroughly and select the painting or picture that you would buy for your friend".*

The experiment instructed subjects to think-out-loud and verbalize everything that went through their minds and comment freely regarding their activity in the web page. The instructions encouraged subjects to focus on navigating the page as you would normally do and to take as much time as necessary to complete the task.

A brief questionnaire at the end of each task evaluated their browsing experience. Usability questions indicated webpage complexity, ease of navigation, and if subjects felt comfortable using the website. Design questions indicated the webpage visual appeal and if subjects liked the webpage design. Interest questions indicated subjects' interest level. In addition, memory recall questions confirmed if subjects read or scanned the webpage. The questionnaire questions were summarized as interest, usability and design scores. Data collected from the questionnaire also included whether the article was read thoroughly or not.

150 data points for Interest Score, Design Score, Usability Score, and Time on task were collected from 28 participants. Each data point was associated with mouse movement activity and observations. Interest score represents explicit interest ratings collected on a five point scale: no interest, low interest, neutral, moderate interest, and high interest. User responses suggested that there was not a clear distinction between degrees of interest, so the scale was later re-coded to appropriately reflect user responses and improve classification accuracy. The new scale (no interest, neutral, and high interest) was used in subsequent analysis. The user mouse behaviors were summarized in categories reflecting the behaviors that occurred in the course navigating a webpage. The categories are mouse over, select text, pause-and-read, think-decide, and scroll.

7.7.3.B Results

Significant correlations between interest, design score, task time and reading condition indicate that interest might be influenced by the visual appeal of a web page, and the time spent reading or scanning a web page (See correlations on Table 1). These correlations support a one-way ANOVA tests showing statistical difference in interest score due to the type of web page type. Graphic intensive web pages were ranked as highly interesting, whereas web pages with links only were ranked the least interesting. These differences indicate that web page type and the elements within a web page might be good predictors of user interest.

Correlations

		INTEREST	EXPLVL1	EXPLVL2	DESIGN	TASKTIME	CONDI_BI
INTEREST	Pearson Correlation	1.000	-.021	.032	.604**	.285**	-.291**
	Sig. (2-tailed)	.	.801	.702	.000	.001	.001
	N	141	141	141	141	136	123
EXPLVL1	Pearson Correlation	-.021	1.000	.769**	.069	-.126	-.157
	Sig. (2-tailed)	.801	.	.000	.414	.142	.081
	N	141	142	142	141	137	124
EXPLVL2	Pearson Correlation	.032	.769**	1.000	.145	-.137	-.177*
	Sig. (2-tailed)	.702	.000	.	.087	.110	.050
	N	141	142	142	141	137	124
DESIGN	Pearson Correlation	.604**	.069	.145	1.000	.132	-.286**
	Sig. (2-tailed)	.000	.414	.087	.	.125	.001
	N	141	141	141	141	136	123
TASKTIME	Pearson Correlation	.285**	-.126	-.137	.132	1.000	-.025
	Sig. (2-tailed)	.001	.142	.110	.125	.	.791
	N	136	137	137	136	137	119
CONDI_BI	Pearson Correlation	-.291**	-.157	-.177*	-.286**	-.025	1.000
	Sig. (2-tailed)	.001	.081	.050	.001	.791	.
	N	123	124	124	123	119	124

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 4 Explicit User ratings under several webpage types. Design score, Task time, and Web page time show significant correlations.

7.7.4 User Interest and Activity Classification

Research on web-analytics has shown that the web offers a vast set of data suitable for pattern recognition algorithms. Pattern recognition algorithms from web-server logs show some promise in detecting and understanding user trends and preferences. Researchers have tried to correlate the time spent on a webpage with a wide range of variables (Claypool 2001).

Our previous studies and experiments investigating people's web browsing behaviors provided the necessary data to create machine learning algorithms that can identify activities from mouse tracks alone. We used supervised learning with an explicit training phase to develop a system that predicts user activities and user interest using algorithms that correlate mouse tracks and implicit metrics to activity and interest. Clustering low level mouse data into a relatively small set of features proved vital in tracking people's behaviors as they occur.

7.7.4.A Design and implementation

The data collection stage previously described provided a detailed record of user web navigating activities. A data integration stage correlated the data and the activity labels and trained an activity classifier using relevant features and activity labels. In addition, in a final learning stage, an activity classifier is able to create a model of the user's activities and predict activities based web-usage observations. Activity labels were used to train and validate the results of the activity recognition algorithms. Activity labeling is a reliable method for labeling people's activities. Direct observation and user input provided qualitative and quantitative measures for computer user experience and interest levels.

7.7.4.A.1 Feature selection/extraction

Transformations from raw mouse activity and web page interactions were used as features to predict several user activities. Features extracted from mouse activity data were mean, variance, energy and entropy for each axis. These features were calculated over 50% overlapping sliding windows, with multiple samples per window. Several feature windows at t + DELTA t were included for analysis, with DELTA = 1, 10, 100 and 500ms values. This allowed the activity recognition algorithm to produce classification outputs at semi-continuous intervals. Mouse activity features were calculated by taking the sum of the squared discrete FFT magnitudes, the entropy of the discrete FFT magnitudes and dividing by the window length.

Extracted webpage interaction features were time spent on webpage, time and number of times on a paragraph, image, link, list, time spent on all elements, and webpage element transitions, such as transitioning from a paragraph to an image and from an image to a paragraph(P-P P-IMG P-A P-DIV P-UL IMG-P IMG-IMG IMG-A). Figure 39 shows activity transitions within several elements on a web page.

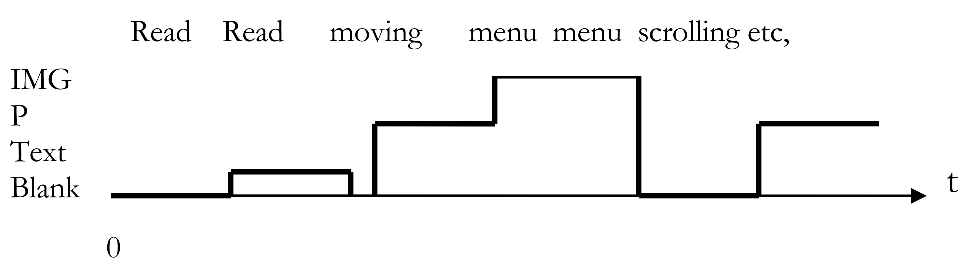


Figure 39 User transitioning within the elements in a webpage (from one element to the other) over a time window.

The Document Oriented Model DOM is supported by most web browsers to store all he elements within a webpage. Our tracking system gathered DOM data through API function calls and created a webpage interaction matrix that summarizes user navigation. This matrix indicates how users interacted with the different elements in a webpage, supporting our goal to use simple computation and universally applicable trackers.

A webpage-element interaction matrix can be used to differentiate user activity and interest within a web page. A simple scenario on an online news article demonstrates this approach. Users thoroughly reading article interact with the areas of the webpage relevant to the article, such as the content section and paragraphs in the webpage. On the other hand, users not interested or just scanning the article can be easily distracted by other elements in the webpage, such as imagery and links to other articles. In the activity user study, subjects were directed to a news article (from the CNN website) and instructed to either read or scan the content. Subjects who found the article interesting concentrated their mouse movement activity on paragraph elements; their interaction matrix shows a high P interaction ratio; in fact all of their activity centered on the article text, see Figure 40-left. The interaction matrix for subjects who found the CNN article uninteresting is comprised of a wider spectrum of interactions—many other elements in the webpage are involved, see Figure 40-right.

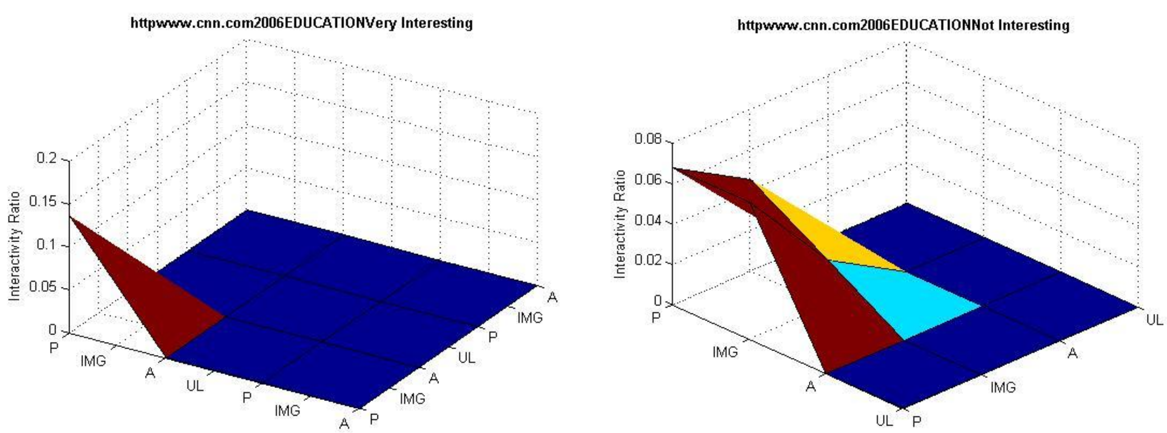


Figure 40 Interactions for CNN pages identified as very interesting (left) and not interesting (right). P= Paragraph A = Link UL = List IMG = Image

7.7.4.B Implementation

The interest detection system implemented a decision tree; one of the most widely used and practical techniques for inductive inference (Mitchell 1997). Decision trees are computationally efficient, and their performance is suitable for real-time recognition (Mitchell 1997). A tree J.45 decision tree was implemented using a Java based pattern recognition tool kit (Witten 1999). The J.45 was selected due to its ability to execute properly on the limited resources offered by web browsers, and due to its robustness to errors. The classifier implemented a pruned tree to avoid over fitting the data and focused on pages with medium to large text bodies in order to improve prediction reliability.

The features that provided the best discriminating power were selected based on the computational power required to analyze them. The features selected are time on text paragraphs, number of times on link, image-paragraph interactions, image-image interactions, and link-list interactions. Interactions features were calculated in ratios; based on the total interactions with other elements. A DELTA t = 500ms best summarized user activity and improved prediction results. Predictions are calculated based on 500ms intervals and are later normalized to obtain a final interest score for each webpage. The classifier was validated using the 10-fold cross-validation method.

7.8 Chapter Summary

This chapter demonstrated the feasibility of using low-level implicit metrics to provide information about people's activities while navigating web pages, using email and responding to Instant Messages. The chapter also described the implementation of several systems that detect user activity, infer concepts surrounding those activities, and predict interest from several websites. Demonstrations presented in this chapter are later used in the next chapter to implement a disruption manager to control interruptions.

The algorithms and tools developed in this chapter have been used to support other research questions, such as the use of mouse tracks for usability and web design (Arroyo 2005), relevant news aggregation from implicit site interactions (Arroyo 2007), a word-based interruption manager (Shawn 2007), and a new research methodology named "recycled-research" (Hockendougal 2007).

The chapter presented several explorations and experiments evaluating mouse tracks and textual information as implicit metrics for understanding disruption. The chapter presented results from these experiments and their application on web browsing and disruption management.

C H A P T E R

8

DISRUPTION MANAGER

8.1 Introduction

The disruption model and framework described in Chapters 4 and 5 are based on existing literature related to information processing, memory, attention and experiments evaluating the effects of interruptions. Therefore it is important to test and demonstrate the model "usability" with respect to practical applications. The framework ecological validity is tested with a disruption manager designed to work on web browsing and instant messaging environments.

This chapter details the design and implementation of a disruption manager designed to mediate interruptions based on lessons learned from our exploratory experiments described in Chapter 3, the disruption model described in Chapter 4, and the framework outlining the factors needed to mediate disruption in computing activities described in Chapter 5. We have focused on using context information regarding people's goals and designed a disruption mediator that supports those goals.

Chapter 7 demonstrated several data sources as implicit metrics. It evaluated using mouse tracks and textual information as virtual sensors for understanding disruption. These explorations provided all the necessary tools for implementing a disruption manager.

The disruption manager uses goal concepts and task context as the main factors in predicting disruption. The manager implements an interruption model, however, its implementation does not attempt to replace computational cognitive models, such as the EPIC and ACT-R (Kieras 1997, and Anderson 1998); which simulate and predict user performance when interacting with computer interfaces.

8.2 Scenarios

The disruption manager utilizes several modules that present complex behaviors that are more easily explained through several scenarios.

8.2.1 Intermediate Layer

A student working on a term project goes to wikipedia to find information about the OPEC. He reads the wikipidia article carefully as he is trying to understand how the OPEC controls gas prices. As he reads, his news reader receives a new RSS feed, however the notification is delayed by the disruption manager as his mouse behavior indicates that he is actively reading the wikipedia webpage. The news feed notification is delivered when he switches to another application, or as he loads a new webpage. At this point the manager assigns a high interest score to the OPEC wiikipedia article based on the student interaction with the webpage, and the concepts associated with this page become part of the user-modeled goals for the given time.

8.2.2 Relevancy and feedback

After reading the wikipedia article, the student decides to read his email and starts responding an email to his sister. As he writes the email, his news reader receives a RSS feed about the OPEC and petroleum price increases. The manager determines that this message might be relevant to one of the user goals and presents the notification when the student submits the email. The news abstract attracts his attention and finds it relevant, so he decides to open the news article webpage and reads it. As he reads, the manager learns that its prediction was appropriate and that the concepts relevant to the article should remain as part of the user goals.

8.2.3 Concepts

The manager has been observing the students for a while now as it has delayed some notifications. The manager slowly builds up new concepts related to travel and vacation based on the student conversations and emails with his friends. The student continues to work on his paper, and as the deadline gets closer, he ignores most of the IM and email notifications; indicating that he is busy on his task. At this time he receives a new email from his friend about travel and the manager adjust the notification transparency. The student notices it and replies to the email within five minutes, interrupting his work, but capitalizing on a great vacation deal.

8.3 Disruption Manager System

Some components in the interruption model are implemented on a disruption mediator designed to balance timing and the amount of interrupting messages people receive while performing their daily computing activities: browsing the web, sending and receiving email, text processing, etc.

The mediator's implementation of the detection stage includes an interruption filter, which captures instant messages (incoming interruptions), interprets if an action is required from the user by extracting normalized verb-subject-object-object frames from the body of the message, classifying the message as an interruption if it carries an associated task, or as a notification otherwise. The detection stage is also responsible for categorizing incoming messages based on the concepts relevant to their content by performing spread activation from the original message concepts. The mediator also maintains contextual awareness by continuously monitoring the topics relevant to the user's activities from a history of opened documents and visited web pages.

The mediator's implementation of the decision stage takes the concepts carried over from the detection stage and computes their weighted contextual-intersection using conceptNet. The topics are then evaluated and compared to the topics relevant to ongoing and past goals. This determines the relationship between ongoing goals and the content of the interrupting message. If relevant, the interruption will be allowed even if the user state indicates the user is currently engaged, thus maximizing the chances for users to accomplish their goals.

The mediator maintains a pool of concepts with constantly changing priority values; similarly to the pool of tasks and hierarchy of goals in the interruption model. Priority values are based on matching concepts from recently accessed web pages and search queries. Items from past activity loose priority over time according to a power function (Anderson 1990). Greater weights are placed on more recent concepts and past information will decline as new concepts are added. Notifications might elicit the user to engage in a new task depending on their content. Thus, concepts associated with notifications are also added to the pool of concepts being tracked. Interrupting messages are treated the same way.

The mediator handles internal interruptions by keeping track of ongoing information retrieval. As users navigate similar web pages, related concepts will become prominent. However, if the user decides to search for new information due to an internal interruption, then priority values will gradually reflect the topic change as if it had been triggered by an interrupting message. The mediator does not emulate the planning component of the mental model decision process, such as preemption, goal activation and resumption and long term memory stores.

8.4 Disruption Manager Implementation

We developed a test bed to evaluate dynamic interruption systems. The test bed allows the examination of the relation between ongoing behaviors, task actions, goals and interruptions.

Applications based on the interruption model are implemented as a three layer architecture system. A low level layer includes implicit low granularity information such as key-strokes and mouse movement activity. An intermediate layer includes the activities and information to which some of the low granularity data can be extracted and summarized, such as reading, switching tasks, paying attention, etc. A top layer or knowledge layer includes the information or concepts relevant to the user goals, see Figure 41.

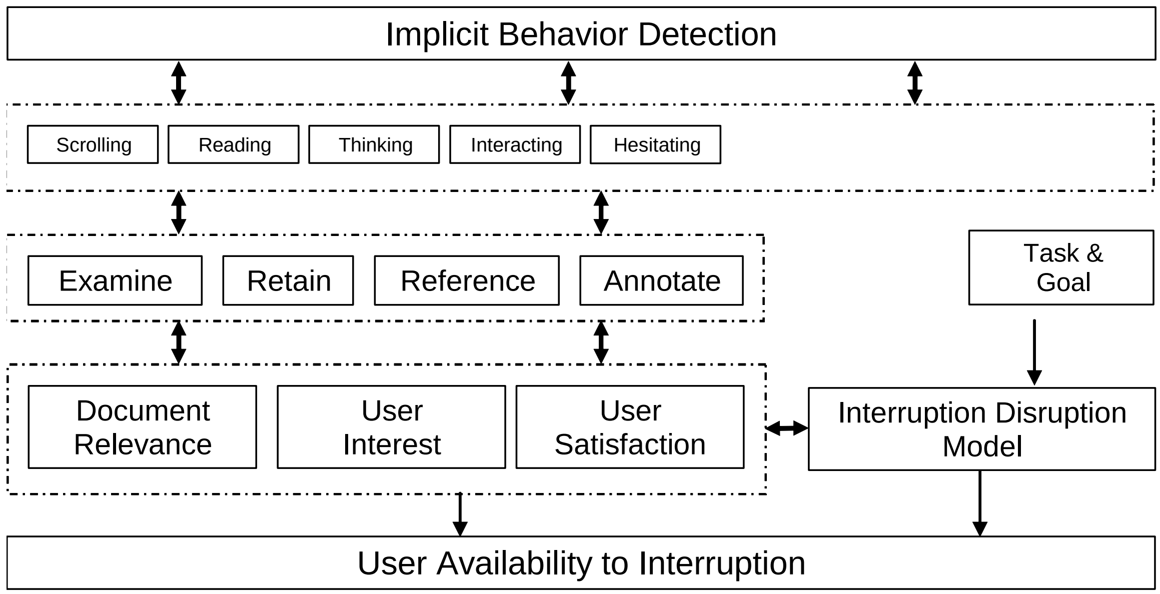


Figure 41 Three-layer architecture: low level information is classified into user activities, and a higher layer represents topics relevant to the user's goals.

The disruption manager monitors the user state (current activity), concepts surrounding the user's goals: history of recently accessed documents, web pages and search queries, the interrupting message relevance to these concepts, and concept priority. The manager then identifies interrupting messages that should be allowed to reach the user immediately or that should be delayed to an appropriate time within task execution.

The disruption manager uses several monitoring modules to track the user state, concepts surrounding the user's goals and interrupting message concepts. The manager analyzes instant messages as they are received, evaluates the appropriateness of the interruption to the user in several contexts based on a model, and then controls the timing and the presentation of that instant message. The system has one module for each context in which the instant message is examined and a decision module that mediates instant message interruptions on multiple Instant Message services based on the evaluations of the context modules. The mediator uses several auxiliary modules for interfacing with the instant message client to both read the instant message content and manipulate the timing and presentation of the instant message.

- Goal level layer
    - Natural Language
- Intermediate Layer
    - Experience Submodule
    - Interest Submodule
    - Reading Submodule
- Low Level
    - Mousetracking Module
    - Task Stage Module (TS Module)
    - Existing Tasks Module (ET Module)

8.4.1 Context Modules

Each Context Module is responsible for evaluating a particular aspect of the interruption, the system, and the user. These modules are derived from aspects of the disruption model. The modules convey their evaluations to the Decision Module as a number, usually indicating the percent appropriateness of showing the instant message at a given moment.

The manager's top level monitoring layer uses Google Desktop and ConceptNet engines as services running on the user's computer. Google Desktop keeps an up-to-date index of files and documents and their contents. ConceptNet is a commonsense knowledgebase with facts from the Open Mind Commonsense corpus (Push 2003). Its concise semantic network contains 200,000 assertions and supports practical textual-reasoning tasks over real-world documents.

8.4.1.A Natural Language Module

The Natural Language Module (NL Module) implements the part of the disruption module concerning the relatedness of the content of the instant message to other documents the user is working with. It uses natural language processing and commonsense reasoning to develop an understanding of the interruption and documents, and attempts to compare the interruption to each document. These comparisons are aggregated into a relevance score, indicating the relevance of the instant message to all documents examined. It also provides the ability to obtain the interpretation of the instant message only, and to compare the interruption to an individual document to determine the relevancy of the interruption.

The NL Module has two major components. The first is responsible for locating files of interest and obtaining the contents of those files. The second component is responsible for natural language processing and commonsense reasoning on data from the first two components.

The first component uses Google Desktop's indexes and caches to access documents on the user's computer. The manager application queries the Google Desktop Engine for recently accessed documents, files of interest (PDF, DOC, PPT, etc), emails, instant messages and web pages and parses them using ConceptNet and a natural language processing engine. Files of interest are simply the files open on the user's computer, as well as recently viewed documents and webpages. The system also obtains a list of open files using VBScript and the Microsoft PsTools library (Microsoft.com). The system then uses Google Desktop (google.com) to locate and read those files. The system also uses Google Desktop to search for recent (viewed in the past hour) webpages in the web cache, and to find and read the documents in the user's My Recent Documents folder. Google Desktop was chosen for its power and speed in searching for files, ability to search based on usage time, and ability to easily obtain the contents of a wide variety of file formats.

The second component uses document-level functions in ConceptNet (text normalization, commonsense-informed part-of speech tagging, semantic recognition, chunking, surface parsing, thematic-role extraction, and pronominal resolution) to extract the verb-subject-object-object frames from recently accessed documents. The entire contents of both the instant message and all of the retrieved documents are individually fed into the MontyLingua (Liu 2004) natural language processing suite. The MontyLingua suite provides both lexical parsing of text and commonsense reasoning through the OpenMind (Singh 2002) commonsense database.

The NL Module extracts from the MontyLingua interpretation key words and concepts in the texts, uses a thesaurus to find possible synonyms for those words and concepts, and then counts the number of times the important words, concepts, and synonyms from the instant message appear in the other documents. The NL module then extracts all the concepts in a document, assigns them saliency weights based on lightweight syntactic cues, and computes their weighted contextual intersection. Concept connections in ConceptNet's semantic network allow the contextual neighborhood around a concept to be found by performing spreading activation radiating outward from a source concept node. The more frequently the number of important words or concepts appear, the more relevant the content of the instant message is likely to be. The output of the NL module is the average number of times a key word or synonym in the instant message appears per sentence in all the searched documents. Preliminary tests I have done have shown the current design to be fairly accurate.

This module allows the manager to summarize text of active documents, identify the documents gist topics, evaluate notifications, capture and classify incoming messages, detect if actions are required from the user, keep track of topics relevant to ongoing and past goals, and determine if incoming interruptions should be presented to the user.

8.4.1.B Moustracking Module

The Mousetracking Module observes mouse usage, and represents the portion of the disruption module concerning the user's depth of involvement in their task activity. This component records the user's mouse movements on a website and reasons about the user interest and activity based on the classifiers and experimental data described in Chapter 7. The Mousetracking Module serves to determine the user interest level in a website and whether the user is reading or scanning a website.

The manager's low level monitoring layer is a proxy-based installed on the user's computer to monitor and categorize mouse movement activity into low granularity behaviors (scrolling, menu, text input, clicking) and user states (reading, deciding, scanning, and waiting). The proxy, a local Apache web server and PHP scripts, fetches web pages requested by the user inserts JavaScript code (see Figure 42). Returned enhanced versions of these web pages have mouse tracking capabilities. The mouse movement tracker filters out very fine movement to reduce data size, speed up classification and remove unnecessary information.

The mouse tracking module outputs data representing the percent interest, and the percent likelyhood the user was closely reading a webpage. The Decision Module uses these heuristics to estimate the depth of user involvement with their current task, with the idea being that the more deeply involved a user is with their current task, the more costly it is to interrupt the user.

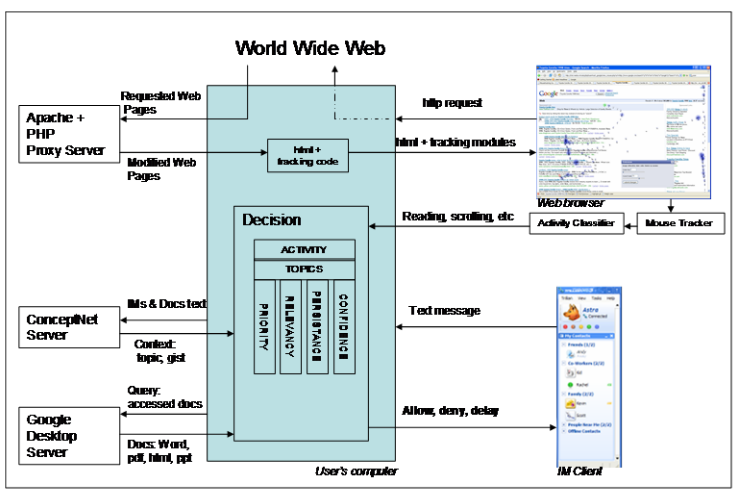


Figure 42 Disruption manager functional diagram. The manager is aided by agents monitoring user activity, document and interruptions topics. Incoming IM can be allowed, denied, or delayed

8.4.1.C Task Stage Module

The Task Stage Module (TS Module) is responsible for determining at what point during some task a user is, which in the disruption model concerns the cognitive load and degree of involvement with the user. It attempts to determine whether the user is at the beginning, in the middle, at the end, or between tasks. It does this by looking for discontinuities or changes in keyboard and mouse usage and windowing behavior. The task stage module looks for significant changes to the number of keystrokes per minute, mouse time per minute, or windowing behavior indicating that the user is changing tasks, or at least subtasks, and thus at those moments interruptions are more likely to be less disruptive.

8.4.1.D Existing Tasks Module

The Existing Tasks (ET) Module attempts to gain an understanding of persisting tasks the user may have, even though they are not currently working on them, and corresponds to the part of the disruption module which determines if interruptions relate to other tasks the user has but may not be currently working on. The ET Module returns the percentage of past tasks an interruption is appropriate for, essentially using some of the other Context Modules to evaluate the interruption with information from the past tasks. This information is used as part of the heuristics for how often the user cares about the interruption topic , and thus includes some information on how significant tasks related to the interruption are to the user.

8.4.2 Decision Logic

The Decision Module is the central component of the disruption manager. Whenever an instant message arrives, the Decision Module determines the appropriateness of that message. It polls all Context Modules for their evaluations of the instant message, and decides how to proceed. Once the disruption manager decides an interruption should be presented, it delays the interruption until an appropriate time in order not to disrupt the ongoing micro-task or activity. Delaying standards are slightly lowered linearly to guarantee that at some point the message will be displayed to the user. If a message has failed to be delivered, it will be automatically shown to the user, regardless of appropriateness.. However, if the interruption is relevant to the user's goal, the manager gives priority to this interruption, and presents it as soon as possible; while minimizing disruption on the ongoing task.

The manager's decision rules are based on findings from interruptions experiments evaluating interruptions relevancy and priority. These findings show that interruptions relevant to topics the user has worked on have the potential to be valuable to the user's goals, and therefore, should be allowed. Thus, the manager limits the number of irrelevant interruptions in order to reduce perceived disruption.

The findings also show that as the ratio of prioritized topics vs. non prioritized topics increases, users are more likely to be disrupted. Therefore, the manager limits interruptions whenever this ratio increases and allows interruptions relevant to prioritized topics whenever confidence values are above predefined thresholds.

The decision module for the disruption manager is implemented as add-on to Trillian™; a fully featured stand-alone chat client that supports AIM, ICQ, MSN, Yahoo Messenger, and IRC. This allows the disruption manager to be easily deployed and integrated into current systems without any burden on the user, such as migrating existing contacts, learning a new interface, or working with an untested client. Furthermore, the Instant Message client provides unique customization functionality, such as, contact message history, and an advanced automation system to trigger events based on anything that happens in the client. This allows the disruption manager to "catch" incoming interruptions and control them. Figure 42 shows the chat client as part of the manager.

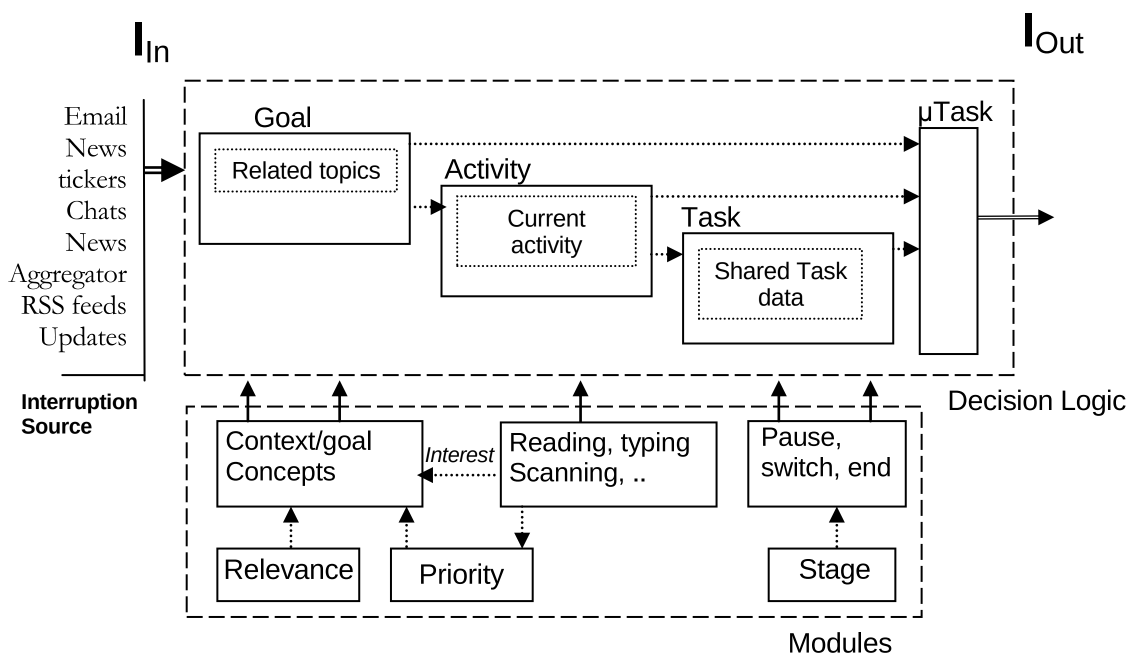


Figure 43 Disruption manager's layered filtering process. The decision Logic is aided by auxiliary modules.

8.5 Evaluation

An experiment was designed to evaluate how effective the disruption manager is in mediating interruptions based on productivity and perceived disruption. Productivity refers to objective metrics designed to evaluate performance for the given task. It is measured using the following metrics: performance, overall goal completion, the time taken to finish an activity, task or goal. Perceived disruption refers to subjective metrics designed to evaluate the user satisfaction for the given tasks and overall goals.

8.5.1 Hypotheses

People under Disruption Manager will have higher performance than people under the No-Manager condition.

People under Disruption Manager will be more efficient in their task than people under the No-Manager condition.

People under Disruption Manager will report less perceived disruption than people under the No-Manager condition.

8.5.2 Experimental Design

The effectiveness of disruption manager was assessed using a between-subjects experimental design with manager and no manager as experimental conditions. Subjects were randomly assigned to one of the two manager conditions: Disruption Manager, and No Mediation. On one condition, interruptions are mediated by the disruption manager, on the other condition interruptions are presented as they arrive. The main dependent variables were performance and perceived disruption. Several other variables were used to confirm the task was performed properly. These variables included task time, number of notifications attended to, time spent on each email, STAI (state trait anxiety) score.

8.5.2.A Task Details

The scenario consisted of customer service and order processing activity for an e-commerce site. The scenario described a typical small business environment where customer service representatives take email orders from several customers and process each order individually trying to satisfy the customer's demands and complete a sale. The task is identical to the customer-based scenario described in the Chapter 6. See the disruption experiments section and Appendix D.1 for details. The scenario explained that customer service representatives obtain a commission based on their sales and instructed subjects to play the role of a customer representative. Adding this role guaranteed that subjects would perform the task to the best of their abilities and encouraged subjects to obtain a bigger profit.

8.5.2.B Disruption Manager

The disruption manager controls email notifications presented based on whether the email is relevant to the ongoing activity and several factors. Figure 43 shows the filtering stages that each interruption must go through before being delivered to the user. The manager allows people to complete the task without unnecessary distractions. That is, relevant Instant Messages are presented (almost) right away so that the subjects can benefit from the Instant Message. On the other hand, irrelevant Instant Messages are delayed until a subtask is finished the. The manager's behavior can be summarized with the following rules:

- Relevant IM are presented after small changes in activity, such as quick task switches, or after finished finding an item, updating values, text entry, etc.
- Irrelevant IM are presented after subjects finish gathering data for one customer, or finish sending email.
- Allow Instant Message notifications if relevant to current email /customer request (active email, document, or webpage).
    - Relevant presented almost immediately.
    - Wait until finished task or task switch.
- Delay Instant Message notifications if relevant or moderately relevant to current email /customer request (active email, document, or webpage)
    - Wait until a task break (Bailey, et al).
    - Only wait for a task switch
- Delay Instant Message notifications if not relevant to current email /customer request (active email, document, or webpage).
    - Wait until email sent.
    - Even if there is a task break.

8.5.2.C Protocol

Each mediator condition was presented in three stages, an introduction stage, a quality stage, and an urgent stage. The introductory stage served to familiarize subjects with the experimental task. This within-subjects condition explored how the task is performed when the task is highly prioritized. The quality and urgent conditions were selected because they exhibited similar traits on the experiment described in chapter 6.2. These highly prioritized tasks were identified as having a larger impact on subjects' Instant Messaging behavior.

40 subjects were randomly assigned to two conditions: Disruption Manager, and No manager. Subjects were first presented with the interface and a walkthrough of the task based on a script previously rehearsed by the experimenter. In order to obtain a consistent response to interruptions, the

walkthrough included an exemplification of potential interruptions and how subjects should deal with them.

A practice session allowed participants to become familiar with the computer-based-test interface, familiar with the content, and familiar with the interrupting messages. The practice session also allowed subjects to identify the benefits from attending to interruptions so that they wouldn't ignore them altogether. The practice session lasted until subjects completed all questions and were satisfied with their answers. On a second practice run, timed sections were introduced in order to introduce this feature and allow subjects to experiment with different navigations techniques.

8.5.2.D Results

The hypothesis regarding performance was confirmed. From the graph in Figure 44, it is clear that mediating interruptions yielded higher performance than without mediation.

Planned comparisons indicate a significant difference on performance based on the manager type $F(1, 37) = 473.92$, $p<.001$. The disruption manager conditions showed 26% performance increase for tasks prioritized by quality and 32.5% performance increase for tasks prioritized by urgency.

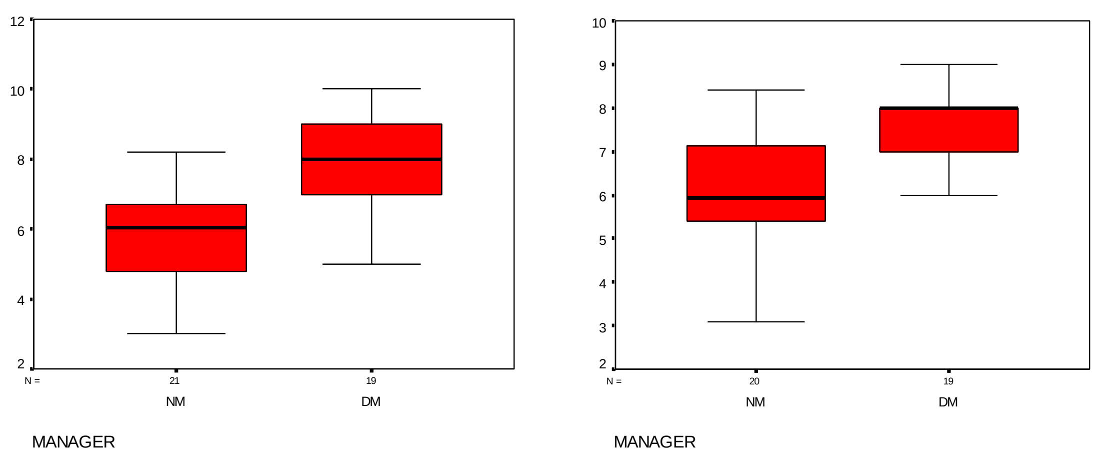


Figure 44 No Manager and Disruption Manager's performance scores for Urgent (left) and Quality (right) conditions.

Additionally to an increase in performance, people were able to share pricing information by replying to Instant Messages, therefore improving overall goal completion (which required participants to share information when possible in order to improve the company's profits).

The hypothesis stating that subjects under the Disruption Manager condition would be more efficient in their task was confirmed. Part of the task included collaborating with other sales associates, thus responding to Instant Messages was also an important part of the task. The ratio of Instant Messages responded was higher for the manager condition. This indicated that the manager did better at presenting interrupting messages (relevant information) at the right time. Participants on the manager condition responded to 58% and 51% of the instant messages received for quality and urgent tasks. Whereas, they only responded to 12% and 8% of the messages on the No-manager condition, see Figure 45.

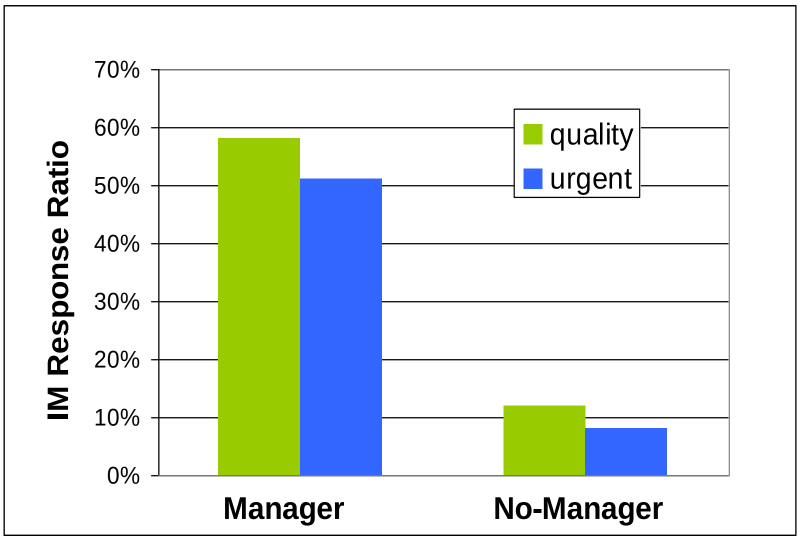


Figure 45 Instant Message Response Ratio. The ratio of IM's responded was higher for the disruption manager,

8.5.2.E Perceived disruption

There was no main effect of manager type in perceived disruption $F(1,37) = .089$ $p=.7$, nor were there any significant contrasts between Quality and Urgent tasks. Thus, our third hypothesis was not confirmed. Both manager categories demonstrated a similar disruptive effect across all task categories as shown in Figure 46. However, our results show a trend towards lower perceived disruption for the disruption manager condition.

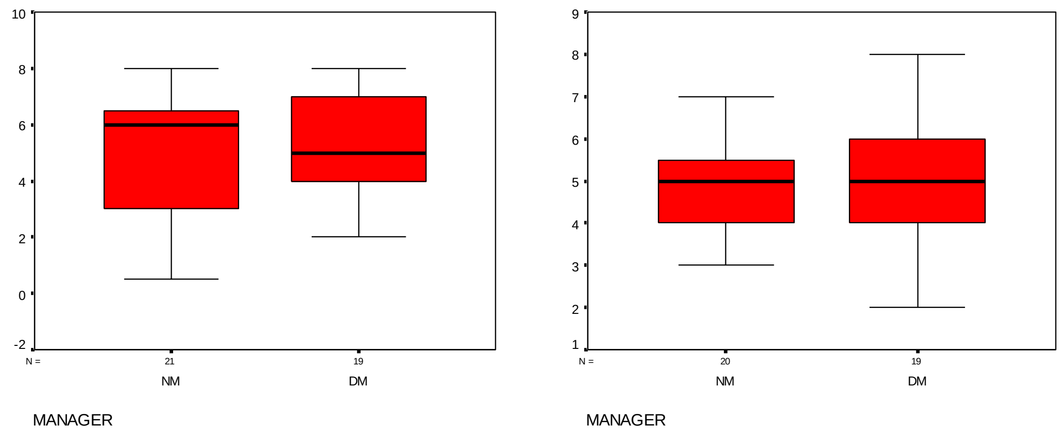


Figure 46 Perceived disruption scores were similar across all conditions and manager type.

8.5.3 Discussion

The results of our experiment demonstrate that computer interfaces are able to manage interruptions and to reduce their disruptive effects. Our results go as far as showing that computer interfaces capable of evaluating incoming interruptions in relation to their benefits to the user's goals and the disruption to the ongoing task can improve performance and overall productivity. Our evaluation show a 26% and 32.5 % performance increase for task prioritized by quality and urgency. We expected that delaying irrelevant interruptions would make them be perceived as less disruptive than they really are. Although not confirmed, our results are still promising since the manager did not interfere with the user goals and did not increase disruption.

8.6 Chapter Summary

A top-down approach and the interruption model were used to develop a disruption manager that controls interruptions on common desktop computing activities, such as web browsing and instant messaging. The manager monitors ongoing behaviors using implicit metrics (virtual sensors) to control possible disruptive outcomes given the user and system state. Virtual sensors provided inferences about the scope of people's goals and tasks. These inferences were generated from domain-independent implicit metrics of interaction (mouse and keyboard behaviors, concepts surrounding the user's goals, interruption relevance, and task priority).

The manager demonstrated that by using simple implicit sensors, it is possible to minimize the disruptive effects of interruptions and increase overall user satisfaction by supporting the user's goals. That is, goals relevant interruptions are supported at the expense of task level disruption. By supporting the user goals, interruptions can be mediated optimizing user satisfaction, work flow, and disruption.

C H A P T E R

9

CONCLUSIONS

There has been a major change in the way people perform their daily activities. People now have access to multiple channels of information and communication, and allow themselves to be constantly interrupted in order to keep track of today's demanding environments. Today's computing activities are performed through multitasking, as technologies enable people to do just that on mobile or desktop environments. These multitasking environments have an increasing number of tasks competing for user's attention. Unfortunately, people have cognitive limitations that make them susceptible to errors when interrupted. People get distracted and often forget details about their main activity due to limited attention span and memory resources that cause interruptions to be disruptive. Disruption of an ongoing activity often negatively affects human performance.

This thesis presents a disruption mediating approach designed to support people's activities, optimize productivity, and minimize the disruptive effects of interruptions. This thesis demonstrates that disruption management is feasible in a wide variety of scenarios, we demonstrated aspects of being able to measure relevant information and interruptions in cars, cell phones, and complex desktop activities.

This work hypothesizes that people's reactions to interruptions are principally affected by goal-oriented strategies. The contextual relationship between interruptions and user goals is an important factor in how interruptions are controlled. The degree to which interruptions are related to the user's goal determines how those interruptions will be received. Extracting and evaluating concepts about the user environment is one way to support the underlying user goals. It is important to recognize that users have several active goals at any given time, and that focusing interruption management to the primary task limits the impact of interruptions on those other secondary goals (see Disruption Model, Chapter 4). Interruptions might be disruptive to the ongoing task, but they might also support a different, perhaps more important goal. This section lists the contributions made in this thesis towards the understanding of disruption in Human Computer Interaction.

1. Context-independent framework for mediating disruption in human computer interaction. This framework includes goal concepts surrounding the user activities, and task context, as provided by implicit sensors.

2. Introduced the differentiation between perceived disruption and task disruption, and the value of focusing on perceived disruption as means for improving user satisfaction and work flow. Demonstrated that perceived disruption takes precedence over performance metrics.

3. An interruption-disruption model guided several experiments investigating some of the factors that play a role in people's decision process regarding interruptions. The work demonstrated and empirically validated that the type of task prioritization plays a significant role in the decision process regarding the acceptance of interruptions.

4. An implemented disruption manager demonstrates the approach of using goal-related concepts, tasks and micro-task as being successful in diminishing negative effects from interruptions. The manager supports goal commitment and task priority.

5. Demonstrated the use of simple implicit sensors to minimize the disruptive effects of interruptions and increase overall user satisfaction. Also demonstrated the feasibility of learning high level user interactions from raw mouse metrics.

9.1 Disruption Management Framework

Interruptions are a growing area of work and researchers have already identified the importance of interruptions in every day computing activities (see Related Work, Chapter 2). Interruptions affect performance and productivity and are very important in the design of human-computer interfaces. Therefore, disruption management is fundamental in achieving optimal work flow and improving user satisfaction.

Existing work has focused on controlling interruptions after having identified areas prone to disruption in very specific tasks. This approach has yielded several guidelines for the design of interrupting and notification systems; unfortunately, they are limited in scope. Our work sets itself apart from previous work in that we analyze interrupting requirements in terms of interruption content and its relationships to higher level user goals and tasks. Our approach focuses on the concepts surrounding the user goals and matches incoming interruptions to those concepts. Our work is based on the premise that people's reactions to interruptions and disruption are principally affected by goal-oriented strategies. Goal and task context serve as important factors for mediating and reasoning about disruption (see Disruption

Model, Chapter 4). We tested the approach in a complex ordering scenario where the user had to pay attention to emails, IM's calculator and filling out forms. When users performed the task carefully, a disruption manager system allowed a for performance increase of 25%. When users performed the task with the goal of having throughput, the system allowed for a performance increase of 31%.

9.2 Perceived Disruption

Our approach places the user as the focal point in the design of interruption-management interfaces, and all efforts are oriented to support the user. Previous work in the area of interruptions has focused mainly on the effects of the type and timing of interruptions on a primary task. While previous work relies on performance metrics on the primary task, we place emphasis on user satisfaction, such as measures of perceived disruption, in addition to performance measures. That is, user goals and motivations take precedence over short-term performance benefits (see Disruption Management Framework, Chapter 5).

9.3 Understanding Disruption Experiments

Our work focused on understanding human disruption and evaluated the effects of interruption relevance to the users' goals, on perceived disruption. We found that interruptions irrelevant to the user-defined goals are perceived as highly disruptive. On the contrary, relevant interruptions are perceived as less disruptive.

Goal and task priority play an important role in the interruption decision process and several experiments evaluated how goal-task context, task priority and task completion level influence on people's reactions and availability to interruptions, and overall task performance. These experiments evaluated four priority conditions during typical multi-tasking activities on desktop computers: no prioritization, prioritized by time, prioritized by quality level, and prioritized by quantity. Our results demonstrated that goal and task priority influences people's susceptibility to interruptions. The results indicate that people recognize the potential benefits of being interrupted and adjust their susceptibility to interruptions during highly prioritized tasks (see Disruption Experiments, Chapter 6). Potentially relevant interruptions are perceived as less disruptive when the user priorities demand attention to detail (quality) or are time-restricted.

9.3.1 Experimental Scenario

An experimental task was developed to study disruption and validate disruption mediation on a realistic and extensible scenario. The scenario was constructed so that it resembles typical multi-tasking activities performed on desktop computers. The scenario reflects commonly computing tasks, such as accessing the Internet or using e-mail. The scenario can be used to evaluate new approaches to disruption management within a common framework.

9.4 Disruption Manager

We proposed a generic approach for mediating disruption. Managing disruption is a complex process that should be addressed at different levels. This thesis demonstrates an example of a computer effectively deciding what to bring to a users attention. We show that taking into account the relative similarity of incoming instant messages and what a person has worked-on can successfully improve disruption management.

The approach uses low-level data to generate inferences about the scope of people's goals and tasks. A disruption manager balances timing and the amount of interrupting messages people receive while performing daily computing activities, such as browsing the web, sending and receiving email, text processing text (see Disruption Manager, Chapter 8). The manager uses semantic similarity, relevance and low level activity levels to decide how to pace Instant Messages (see Implicit Metrics, Chapter 7).

The disruption manager shows that using simple implicit sensors can minimize the disruptive effects of interruptions and increase overall user satisfaction. The disruption manager increased performance by over 25% for conditions of urgent goals and for conditions of careful outcomes. In general, disruption managers have the potential to significantly impact people's lives in positive ways by improving performance and increasing productivity.

9.5 Domain-independent Implicit Metrics

Unlike existing approaches, which are dependent and limited to specific domains, our work focuses on using contextual implicit information regarding people's goals (see Implicit Metrics of Attention, Chapter 7). We demonstrated that concepts representing the user activities and mouse movement activity are reliable and accurate implicit metrics to control user disruption. Domain-independent text streams provide insight onto the user goals, while mouse movements illustrate hidden behaviors reflecting user's attention.

9.5.1 Concepts as implicit metrics

We show that taking into account the relative similarity of incoming instant messages and matching what a person is working on can successfully improve disruption management. We demonstrated that a program can draw inferences about a person's goals by extracting goal-related concepts from text streams, such as instant messages, emails, and documents to determine if the content of monitored interruptions is relevant to a person's work..

9.5.2 Mousetracking as implicit metrics

This thesis validated the use of mouse tracking as a valid tool for understanding users' attention and browsing behaviors. We evaluated the relative value and tradeoffs between mouse monitoring by replicating and extending results from eye-tracking research. Our investigations show that 86% of the people studied used their mouse in an identifiable pattern. Mousetracking also demonstrate the advantages of using unobtrusive implicit tracking;

Several classifiers based on implicit metrics create a model of the user's activities and predict activities based web-usage observations. This model brings high level modeling into the arena of interruption modulation. The model differentiates scrolling, reading, thinking, interacting and hesitating as user actions. It evaluates opportunities for predicting user interest based solely on mouse tracks.

9.6 Anticipated Impact

In most social settings, it is possible to structure technologies in such a way that they control interruptions. Whether closing a door to delimit the space and keeping people outside an office (Yao 2000), or, in more modern times, blocking people on your IM list, these technologies reflect people's availability to interruptions. This work will take that a step forward, instead of passive and physical mediators, we provide active mediators that aim to interpret and recognize the value of communication. These new systems accommodate for different and appropriate ways of handling different kinds of information.

Within the next ten years, we believe that every piece of software will have some element of intelligence that helps it decide when accepting or rejecting information is appropriate. This will have profound influence on social dynamics, the ability for people to accomplish their work, tasks, their homework tasks, their social responsibilities, and even their personal goals.

9.7 Future Work

As computers become more and more powerful, it will be possible to understand more about the user needs and goals (as well as acquiring deep understanding of incoming interruptions). This work has demonstrated that by gaining insight into the user's goals, it is possible to improve the user experience and minimize disruptions. Improved accuracy in matching interruptions to the user goals will provide even further benefit, such as using commonsense knowledge to identify situations where interruptions should be avoided.

Deployment of disruption management systems on large organizations would provide a measurable impact on productivity and user satisfaction. A preliminary study has already demonstrated that people recognize the benefits from a disruption manager and that they are willing to use such a system for extended periods (Sullivan 2007). Although our current system learns from users, further investigation over extended periods, such as longitudinal studies are necessary in order to improve the system's adaptability and feedback customization

The next step in disruption management is the use of a similar approach for every day objects. Considerate artifacts is now part of a new emerging field trying to influence the design of everyday objects to minimize their disruption effects and support effortless interactions. For instance, cellular telephones might automatically switch to vibration and adjust the ringing tone if they are on someone's hands.

REFERENCES

Adams, M.J. & Pew, R.W. (1990). Situational Awareness in the Commercial Aircraft Cockpit: a Cognitive Perspective, *IEEE/AIAA/NASA 9th Digital Avionics SystemsConference,* Institute of Electrical and Electronics Engineers, New York, 519-524.

Allport, A. (1980). Patterns and actions: Cognitive mechanisms are content-specific. In G. L. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 26-64). London: Routledge & Kegan Paul.

Altmann E. M., & Gray W. D. (1998). Pervasive episodic memory: Evidence from a control-of-attention paradigm. In M. A. Gernsbacher & S. J. Derry (Eds.), *Twentieth Annual Conference of the Cognitive Science Society* (pp. 42-47) Hillsdale: Erlbaum.

Altmann E. M. & Gray W. D. (2000). Managing attention by preparing to forget. In *Proceedings of the IEA 2000/HFES 2000 Congress*. (pp. 152-155) Santa Monica: Human Factors and Ergonomics Society.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, 51, 355-365.

Anderson, J. R. & Liebiere. C. (1998). *The Atomic Components of Thought*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Aronson, E. & O'Leary, M. (1982). The relative effectiveness of models and prompts on energy conservation: A field experiment in a shower room. *Journal of Environmental Systems*, 12 (3), 219-224.

Arroyo E., Selker T. and Stouffs A., "Interruptions as Multimodal Outputs: Which are the Less Disruptive?," IEEE International Conference on Multimodal Interfaces (ICMI'02), Pittsburgh, PA, pp. 479-483, October 2002.

Arroyo, E., and Selker, T. (2003). Self-adaptive multimodal-interruption interfaces. In *Proceedings International Conference on Intelligent User Interfaces* (IUI'03).

Atkinson, R.L. & Shiffrin. R.M. (1968). Human memory: a proposed system and its control processes. In K.W. Spence & J, T, Spence, (Eds.) *The psychology of learning and motivation: advances in research and theory*, Vol. 2. New York: Academic.

Bailey, B. P., Konstan, J. A. & Carlis, J. V. (2000). The effect of interruptions on task performance, Annoyance, and Anxiety in the User Interface. In *IEEE International Conference on Systems, Man, and Cybernetics*.

Bailey B. P., Konstan J. A. & Carlis J. V. (2000) Adjusting windows: Balancing information awareness with intrusion. In *Proceedings of the 6th Conference on Human Factors and the Web* (HFWeb 2000-B).

Baddeley, Alan (1996). Exploring the central executive, *Quarterly Journal of Experimental Psychology: Human Experimental Psychology [Special Issue: Working Memory*], 49 (1), 5-28.

Barnard, P. & May, J. (1999). Representing cognitive activity in complex tasks. *Human-Computer Interaction*, 14, 93–158.

Barnard, Philip & May, Jon. (2000). Systems, interactions, and macrotheory. *ACM Transactions On Computer-Human Interaction*, 7(2), 222–262.

Chen, M., Anderson, J. R., & Sohn, M. (2001). What Can a Mouse Cursor Tell Us More? Correlation of Eye/mouse Movements on Web Browsing. *Ext. Abstracts CHI 2001*, ACM Press .

Chewar C. M., McCrickard D. Scott, & Sutcliffe Alistair G. (2004). "Unpacking Critical Parameters for Interface Design: Evaluating Notification Systems with the IRC Framework." In Proceedings of the 2004 Conference on Designing Interactive Systems (DIS '04), Cambridge MA, pp. 279-288.

Chong L. J., Chewar C. M., & McCrickard D. S. (2004). "Image is Everything: Advancing HCI Knowledge and Interface Design Using the System Image." In *Proceedings of the ACM Southeast Conference (ACMSE '05)*, Kennesaw GA, March 2004, pp. 2-376 - 2-381.

Cialdini, R. (2001). The science of persuasion. *Scientific American,* 284, 76-81.

Claypool, M., Le, P., Waseda, M., & Brown, D. (2001). Implicit interest indicators. In Proceedings of the 6'4 International Conference on Intelligent User Interfaces (IUI '01), USA, 33-40.

Cohen, S. (1980). After effects of stress on human performance and social behavior: A review of research and theory. *Psychological Bulletin*, 88, 82-108.

ComScore Worldwide Online Universe Estimate Based on the World's Largest, Most Representative Sample and Most Robust Methodology, (2006, May 04) posted to http://www.technologynewsdaily.com/

Covey, S. R. (1989). *The Seven Habits of Highly Effective People*. New York: Simon and Schuster, Inc.

Cutrell, E., Czerwinski, M. & Horvitz, E. (2001). Notification, Disruption and Memory: Effects of Messaging Interruptions on Memory and Performance. In Hirose, Michitaka (Ed.), Human-Computer Interaction INTERACT '01, (Tokyo, July 9-13), IOS Press (for IFIP), 263-269.

Czerwinski, M., Cutrell, E. & Horvitz, E. (2000-A). Instant Messaging: Effects of Relevance and Time. In S. Turner, P. Turner (Eds), People and Computers XIV: Proceedings of HCI 2000, Vol. 2, British Computer Society, 71-76.

Czerwinski, M., Cutrell, E. & Horvitz, E. (2000). Instant Messaging and Interruption: Influence of Task Type on Performance. In OZCHI 2000 Conference Proceedings. Eds: Paris, C., Ozkan, N., Howard, S. and Lu, S., 2000-B, 356-361.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions, Proceedings of the SIGCHI conference on Human factors in computing systems, p.175-182, Vienna, Austria

DeBell, M., & Chapman, C. (2006). *Computer and Internet Use by Students in 2003 (NCES 2006–065).* U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Enquiro, Did-it and Eyetools uncover search's Golden Triangle, 2005. http://www.enquiro.com/eye-tracking-pr.asp.

Etgen, M. & Cantor, J. (1999). What Does Getting WET (Web Event-Logging Tool) Mean for Web Usability? Proc. HFWEB.

Fabian A., Felton D., Grant M., Montabert C., Pious K., Rashidi N., et. al. (2004). "Designing the Claims Reuse Library: Validating Classification Methods for Notification Systems." In *Proceedings of the ACM Southeast Conference (ACMSE '04),* Huntsville AL, April 2004, pp 357-362.

Fogarty, J., Ko, A. J., Aung, H. H., Golden, E., Tang K. P., & Hudson, S. E. (2005).Examining task engagement in sensor-based statistical models of human interruptibility, Proceedings of the SIGCHI conference on Human factors in computing systems, Portland, Oregon, USA

Fogg, B.J. (2003). *Persuasive Technology: Using Computers to Change What we Think and Do*. San Francisco:Morgan Kaufmann.

Garcia-Ogueta, M. I. (1993). Internal attentional switching: Effects of predictability, complexity and practice. *Acta Psychologica*, 83, 13-32.

Gievska, S., & Sibert, J. (2005). Examining the Qualitative Gains of Mediating Interruptions during HCI. In the Proc. of HCI 2005.

Gibson, J.J.a.C., L.E. (1998). A theoretical field analysis of automobile driving. *The American Journal of Psychology*, 11, 453- 447.

Gillie,T. & Broadbent, D. (1989).What makes Interruptions Disruptive? A study of length, Similarity and Complexity. *Psychological Research*, 50, 43-250.

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N. & Wichansky, A. M. (2002). Eye Tracking in Web Search Tasks: Design Implications. Proc. ETRA p.p. 51-58.

Google Desktop: http://desktop.google.com/

Greenberg S. & Witten I.H. (1985). Adaptive Personalized Interfaces: A question of viability. *Behavior and Information Technology*, 4, 31-45.

Groff, B. D., Baron, R. S., & Moore, D. L. (1983). Distraction, attentional conflict, and drivelike behavior. *Journal of Experimental School Psychology*, 19, 359-380.

Hess, S. M., Detweiler, M.C. Training to reduce the disruptive effects on interruptions. In Proceedings on the Hunan Factors and Ergonomics Society 38th Annual Meeting, 1994, 1173-1177.

Hseih-Yee, I. (2000). Research on Web search behavior. *Library & Information Science Research*, 23, 1–19.

Hong, J.I., Heer, J., Waterson, S. & Landay, J.A. (2001). WebQuilt: A Proxy-based Approach to Remote Web Usability Testing. *ACM Transactions on Information Systems* 19, 3.

Hochheiser, H. & Shneiderman, B. Understanding Patters of User Visits to Web Sites: Interactive Starfield Visualizations of WWW Log Data. Technical Report, University of Maryland, College Park, MD, USA, 1999.

Horvitz E. & Apacible J.(2003). Learning and reasoning about interruption. In *Proceedings of the Fifth International Conference on Multimodal Interfaces*, Vancouver, BC, Canada.

Horvitz, E., Jacobs, A., & Hovel, D. Attention-Sensitive Alerting. In *Proceedings of UAI '99, Conference on Uncertainty and Artificial Intelligence*, Stockholm, Sweden, July 1999. Morgan Kaufmann: San Francisco. pp. 305-313.

Horvitz E., Kadie C., Paek T. & Hovel D. (2003). Models of attention in computing and communication: From principles to applications. In *Communications of ACM*, 46 (3), 52-59.

Hutton K.A., S.C.G., Harper D.N. & Hunt M. (2001). Modifying Driver Behaviour with Passenger Feedback. *Transportation Research Part F: Traffic Psychology and Behaviour*, 4 :4, 257-269.

Intille S.S., Kukla, C., Farzanfar R., and Bakr W. (2003), Just-in-Time Technology to Encourage Incremental, Dietary Behavior Change. in AMIA 2003 Symposium.

Iqbal, Shamsi T. and Bailey, Brian P. (2005): Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In *Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems* 2005. pp. 1489-1492.

Ishii, H., Wisneski, C., et al. (1998) AmbientROOM: Integrating Ambient Media with Architectural Space. *In ACM Conference in Computer Human Interaction.*

Jacko, J. A. (1997). An empirical assessment of task complexity for computerized menu systems. *International Journal of Cognitive Ergonomics*, 1, 137-148.

Jackson T. W., Dawson R. J. & Wilson D. (2001). The cost of email interruption. *Journal of Systems and Information Technology*, 5 (1), 81-92.

Jagannathan, V., Dodhiawala, R., & Baum, L. S. (1989). Blackboard Architectures and Applications: Perspectives in Artificial Intelligence, volume 3. Academic Press, Boston.

Jones, M.Y., Stanaland, A.J., & Gelb, B.D. (1998). Beefcake and Cheesecake: Insights for Advertisers. *Journal of Advertising,* 27, 2 , 33-51.

Josephson, S. and Holmes, M. E. Visual Attention to Repeated Internet Images: Testing the Scanpath Theory on the World Wide Web. Proc. ETRA (2002), 43-49.

Kieras, D. E. and Meyer, D. E.. An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Human-Computer Interaction, 12(4):391–438, 1997

Kahneman, D. & Tversky, A. (1973) On the psychology of prediction. *Psychology Review*, 80, 237-251.

Kreifeldt, J.G. and McCarthey, M.E. Interruption as a Test of the User-Computer Interface. In Proceedings of the 17th Annual Conference on Manual Control, 1981, 655-67.

Latorella, K., "Effects of Modality on Interrupted Flight Deck Performance: Implications for Data Link," Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Chicago, IL, October 1998.

Lieberman, H., Selker T., Out of context: computer systems that adapt to, and learn from, context, IBM Systems Journal, v.39 n.3-4, p.617-632, July 2000

Liu, H. & Singh, P. (2004) ConceptNet: A Practical Commonsense Reasoning Toolkit. BT Technology Journal, Volume 22, forthcoming issue. Kluwer Academic Publishers

Locke, E. A., & Latham, G. P. (2002). Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-year Odyssey. *American Psychologist*, 57, 705-717.

Lockerd, A.L. (2002). DriftCatcher: Understanding Implicit Social Context in Electronic Communication. MIT Master's Thesis, September 2002.

Maes, P. Agents that Reduce Work and Information Overload. In *Communications of the ACM*, Vol. 37, No.7, July 1994, pp. 31-40

Maglio, P.P. & Campbell, C.S. Tradeoffs in displaying peripheral information. In Proceedings of the CHI 2000 Conference on Human Factors in Computing Systems, 2000.

McCrickard D. Scott, & Chewar C. M. "Attuning Notification Design to User Goals and Attention Costs." Communications of the ACM , 46 (3), March 2003, pp. 67-72.

McFarlane D. C. Coordinating the interruption of people in human-computer interaction. A. Sasse and C. Johnson (Eds.), *Proceedings of Human-Computer Interaction (INTERACT'99)*, IOS Press, IFIP, 1999, pp. 295-303.

McFarlane D. C. & Latorella K. A. (2002). The scope and importance of human interruption in human-computer interaction design. *Human-Computer Interaction*, 17 (1), 1-61.

Meyer, D. E., Evans, J. E., Lauber, E. J., Rubinstein, J., Gmeindl, L., Junck, L., et. al. (1997, March). Activation of brain mechanisms for executive mental processes in cognitive task switching.. Cognitive Neuroscience Society, Boston, MA.

Meyers-Levy, J. & Maheswaran, D. (1991). Exploring Differences in Males' and Females' Processing Strategy. *Journal of Consumer Research,* 18, 63-70.

Meiran, N., Chorev,Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology*,41,211- 253.

Michon, J.A. (Ed.)(1993). *Generic Intelligent Driver Support*. London: Taylor & Francis.

Microsoft. PsTools http://www.microsoft.com/technet/sysinternals/utilities/pstools.mspx
Miller, G. A., Galanter E., and Pribram, K. H. Plans and the structure of behavior. London: Holt, Rinehart and Winston, (1960).

Mitchell, T. (1997) Decision Tree Learning, in Machine Learning, McGraw-Hill, pp. 52-78.

Mueller, F. and Lockerd, A. Cheese: Tracking Mouse Movement Activity on Websites, a Tool for User Modeling. Ext. Abstracts CHI 2001, ACM Press (2001)

Münch, S. & Dillmann, Rüdiger, Haptic output in multimodal user interfaces. (1997). In *Proceedings in the ACM International Conference on Intelligent User Interfaces*, 105-112.

Nachshon, M., Modeling cognitive control in task-switching. Journal of Psychological Research v63 n.3, p. 234-249, August 2000

Nagel, Kristine S, Hudson, James M., Abowd, Gregory D. Predictors of availability in home life context-mediated communication, Proceedings of the 2004 ACM conference on Computer supported cooperative work, November 06-10, 2004, Chicago, Illinois, USA

"Nielsen Ratings Press Release", March 28 2006, www.nielsen-netratings.com/pr/pr_060330.pdf

Oviat, S. L. & Cohen, P.R. (2000). What comes naturally. In *Communications of the ACM*, 45-53.

Pan, B., Hembrooke, H., Gay, G., Granka, L., Feusner, M. & Newman, J. The Determinants of Web Page Viewing Behavior: An Eye Tracking Study. Proc. ETRA (2004).

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. (1959).*Journal of Experimental Psychology*, 58, 193-198.

Poynter Institute. Stanford Poynter Eyetrack Project: Study of Reading of On-Line News Site. http://www.poynter.org/eyetrack2000/Project2000

Pompei J., S.T., Buckley S. and Kemp J., An Automobile-Integrated System for Assessing and Reacting to Driver Cognitive Load. in IEEE/SAE Convergence, (2002), 411-416.

Ramstein, C, Arcand, J.F., & Deveault, M. Adaptive User Interfaces with Force Feedback. (1996). In *Proceedings of the ACM Computer Human Interaction*, 406-408.

Rasmussen, J. 1986. Information processing and human-machine interaction. An approach to cognitive engineering. New York, North-Holland, 215 p.

Rayner, K. Eye Movements in Reading and Information Processing: 20 years of Research. (1998). *Psychological Bulletin,* 124, 372-422.

Russell W.D., Dzewaltowski, D.A., and Ryan G.J. The effectiveness of a point-of-decision prompt in deterring sedentary behavior. American Journal of Health Promotion: AJHP.1999

Rubinstein, J., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 763–797.

Salvucei, D. (1999). Inferring Intent in Eye-Based Interfaces: Tracing Eye Movements with Process Models. In Human Factors in Computing Systems: Proceedings of CH199 (pp. 15-20). New York: Addison Wesley.

Scholtz, Laskowski, & Downey. Developing Usability Tools and Techniques for Designing and Testing Web Site. Proc. HFWEB (1998).

Selker, Ted. COACH: a Teaching Agent That Learns. Association for Computing Machinery. Communications of the ACM. New York: Jul 1994. Vol. 37, Iss. 7; p. 92.

Selker, T. COgnitive Adaptive Computer Help (COACH): A Case Study. Advances in Computers, Vol. 47, Academic Press, New York (1998) 69-140.

Selker T., Burlesson W. Context-aware design and interaction in computer systems, IBM Systems Journal, v.39 n.3-4, p.880-891, July 2000

Sharon, T., Selker, T., Wagner L. and Frank, A.J. CarCoach: a generalized layered architecture for educational car systems, in Proc. of the IEEE Int. Conf. on Software - Science, Technology & Engineering, Washington, DC, 2005, pp. 13–22

Singh, P., Lin, T., Mueller, E. T., Lim, G., Perkins, T., & Zhu, W. L. (2002). Open Mind Common Sense: Knowledge acquisition from the general public. *Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*. Lecture Notes in Computer Science (Volume 2519). Heidelberg: Springer-Verlag.

Skinner, B.F, (1991). *The Behavior of Organisms: An Experimental Analysis*. New York, B. F. Skinner Foundation.

Spivey, M. J., Tyler, M., Richardson, D. C., & Young, E. Eye Movements during Comprehension of Spoken Scene Descriptions. Proc. CSS (2000).

Sternberg, S. (1969).The discovery of processing stages: extensions of Donders' method. In: Koster WG, editor. Attention and performance. Part II. Amsterdam: Elsevier Science;. p. 276–310.

Sullivan, S. Limiting Interruption Disruptiveness through Task Analysis. Master's Thesis, MIT Media Lab 2007c
.
SuwatanaPongched, P. A More Complex Model of Relevancy in Interruptions. 2003

Tarasewich, P. and Fillion, S. Discount Eye Tracking: The Enhanced Restricted Focus Viewer. Proc. AMCIS (2004).

Van Bergen, A. (1968). *Task interruption*. Amsterdam: North-Holland Publishing Co.

Ware, C., Bonner, J., Knight, W. & Cater, R. (1992). Moving Icons as a Human Interrupt, *International Journal of Human-Computer Interaction*, v.4, pp. 341-348.

Wickens C. D. and Hollands J. G. Engineering Psychology and Human Performance. PrenticeHall,Upper SaddleRiver, NJ,thirdedition, 2000.

Wisneski, C., Ishii, H., & Dahley, A. (1998). Ambient Displays: Turning Architectural Space into an Interface between People and Digital Information. In *International Workshop on Cooperative Buildings*.

Witten, I. & Frank, E. (1999). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann.

Yan, Hao, Selker, Ted. Context-aware office assistant. *Proceedings of the 5th international conference on Intelligent user interfaces*, p.276-279, January 09-12, 2000, New Orleans, Louisiana, United States

Zeigarnik, B. Das behalten erledigter und unerledigter handlungen. Psychologische Forschung, 1927, pp.1–85.

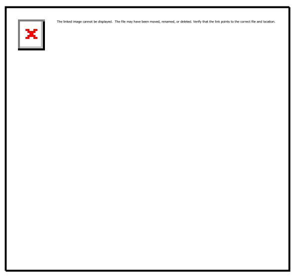
Zijlstra F. R. H., Roe R. A., Leonova A. B. & Krediet I. (1999). Temporal factors in mental work: Effects of interrupted activities, *Journal of Occupational and Organizational Psychology*, 72, 163-185.

APPENDIX

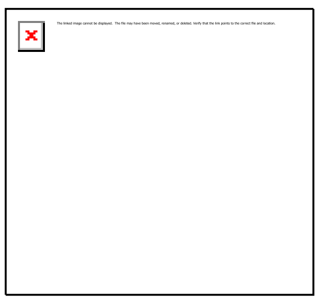
A

ONLINE INTERRUPTING USER STUDY

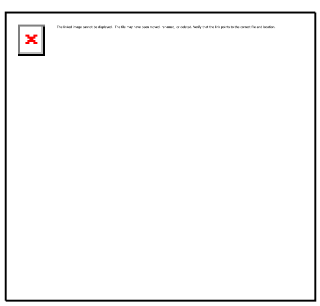
A.1 SUBTASK STEPS REQUIRED TO NAVIGATE EACH WEBPAGE



1	Decide-Commit	Decide to use search engine to find website
2	Enter data	Type search engine URL
3	Wait	Wait for webpage to load
4	Evaluate	Look for text entry box



1	Deicide-Commit	Decide which website to search for, i.e cheaptickets
2	Enter Data	Type and submit search query
3	Wait	Wait for webpage to load
4	Evaluate	Evaluate search results and find desired link



1	Decide-Commit	Decide flight details info to search for
2	Enter data	Enter flight or trip details
3	Wait	Wait for webpage to load
4	Evaluate	Evaluate available flights

1	Decide-Commit	Decide which flight to book
2	Enter Data	Click on flight
3	Wait	
4	Evaluate	Review flight details

A.2 SAMPLE INTERRUPTING RATIONALE RESPONSES

While the map loads, the user must pause for a moment. Once the map is loaded, the user is unlikely to forget what he/she was trying to do after the interruption is over.

User has yet to start any in depth search. I can interrupt them now without stopping any serious data mining.

User has finished typing for the moment. User is waiting for the webpage to load and has a free moment. The user can be interrupted at this time and be able to quickly resume looking at the results when they load.

This person had just finished with one page and was waiting for another to load.

This is the time where they're about to enter the item they're searching for. If they already found relevent information, you shouldn't bother them afterwards.

This is an okay time to interrupt because the person has found their location already so they don't need to concentrate on getting the right address.

There seems to be a pause in the browsing, so the person can look away for a second to pay attention to me.

The user is stationed at google search with no current search started. They may have a search in mind, but they are not deeply involved in the process just yet...

The user has finished typing in the search parameters. The user has obtained the results from the search and is now at a point where, if interrupted, the user can easily resume his activities.

The search is already being interrupted

The person has not yet begun their task.

The person had stopped typing and was waiting for a new page to load.

The page hasn't loaded, so you could work in an opening in a slow, mysterious voice: "So, where are you going?" (awkward obligatory pause) "Can you send me the excel file?"

Since the person pressed search, I know that expedia goes to a page that tells you to wait till they find the info about the trip, thus the person won't be doing anything but wait so its okay to interrupt

same as before, can be interrupted if waiting for another site to open *route was generated*

Person is about to do something that is computationally intensive

options are given, pause can be taken before purchasing ticket

Just got the result from mapquest, not concentrating at the map yet.

Here, the user has just finished pushing the submit button and is waiting for the results. Any change in the screen is expected, and he will not be rudely shocked by any interruption.

He has hesitated for a while, and so it means he is going to stay on this page with his attention.

Entering data has ended and new phase begins.

During this period, he's waiting for the web-search result at least for a few seconds. He's not in the middle of something.

At this time, the person has not decided to select anything yet. There is still time to redirect the person to do something else, before he clicks what he wants to do and proceed.

A P P E N D I X

B

CELLPHONE BASED INTERRUPTIOS EXPERIMENT

B.1 CELL PHONE-BASED INTERRUPTIONS QUESTIONNAIRE

Today's date (MM - DD - YYYY): __ __ - ____ -- _________
Name: ____
Last name: ____
Age: _____
Gender: M F
E-mail: ____
Occupation: ____
High school name: ________________________________
University name: ________________________________
Do you own a cell phone? Yes No
If yes, how long have you owned one? ________

Have you used a cell phone while driving? Yes No

If yes, How many times? _____

What' the main purpose or used for your cell phone? (check as many as needed)

Personal Business Emergency PDA replacement main phone line

Others____

Cell phone features that you commonly use: (check as many as needed)

SMS IM Multimedia SMS email Calendar Alarm Clock

Others____

Activities you do using your cell phone : (check as many as needed)

Taking pictures listening to music, web browsing taking notes

Others____

Commonly used features on your cell phone: (check as many as needed)

Conference calls two-way radio communication speaker phone

Others____

Commonly used accessories with your cell phone: (check as many as needed)

Bluetoth headsets hands-free headsets speaker phone Camera

Others ____

Complete the sentence: *ask not what your country can do for you; ask what can you …*

____

Complete the sentence: *That's one small step for a man, one giant leap*…….

_________________________________-

Please read the following instructions and scenario carefully. Contact the experimenter if you have any questions.

You just found an antique chair at an online auction site. You know that it would sale for 300 on the market and are very interested in buying it. Its current bidding price is 5$ and you know it is a good deal. Paying 100$ for it would be a reasonable price, but you could pay up to 150$, which is your maximum allocated money. Your goal is to make the right decisions to get the chair and maximize your profit. A cell phone based software will present several messages through the duration of the experiment that will provide information about the online auction. After the messages are presented, the cell phone application will ask you to rate how disruptive the message was to your ongoing activity. You can rate the interruption in a scale from 1 to 5 ("Not at All", "Not Really", "Somehow", "Mostly", and "Extremely"). Incoming message will be signaled with a soft chimes sound and gradually increase in volume for about 30 seconds.

Please DO NOT TURN the page until Instructed

B.2 SCENARIO AND INTERRUPTING QUESTIONS

These questions are part of a scenario where subjects are told that one of their goals is to win an online (real time) auction.

Scenario (short description):

You just found an antique chair at an online auction site. You know that it would sale for 300 on the market and are very interested in buying it. Its current bidding price is 5$ and you know it is a good deal. Paying 100$ for it would be a reasonable price, but you could pay up to 150$, which is your maximum allocated money. Your goal is to make the right decisions to get the chair and maximize your profit.

Practice Messages

First two interruptions used to get subjects familiar with the notification and with the interface and actions needed to process an interruption:

1. ***This is an interruption:*** *Accept takes you to this screen to find out more about the interruption. Reject dismisses the interruption.*
2. ***This is an interruption requesting an action:*** *Read the interruption description carefully before deciding/selecting the appropriate option.*
    a. *No action, Another action. Another action, Another Action*

Related Messages

1. ***Entered bidding room:*** *You have entered an auction to bid on "antique chair". The initial bid was 20$. What do you want to do?*
    a. *No Action. Bid +10$, Bid +20$, Exit Auction*
2. ***Bid Increase:*** *The bid on "antique chair" item has been increased to 50$. What do you want to do?*
    a. *No action, Bid +10$ Bid +20$ Exit Auction*
3. ***Auction bid increased:*** *The bid on "antique chair" item has been increased to 80$. What do you want to do?*
    a. *No action, Bid +10$, Bid +20$, Exit Auction*
4. ***Reserve price met:*** *The reserve price for "antique chair" item has been met at 100$. What do you want to do?*
    a. *No Action, Bid 10$, Bid 20$, Exit Auction*
5. ***Similar Auction Item found on sale:*** *A similar "antique chair" item went on sale at an online store for $80. What do you want to do?*
    a. *No action, Track item, Place on hold, Buy now*

Unrelated Messages

1. ***World News: UAE firm to transfer port operations to 'U.S. entity'.*** *United Arab Emirates-owned DP World said Thursday it would transfer its operations of American ports to a "U.S. entity" after congressional leaders reportedly told President Bush that the firm's takeover deal was essentially dead on Capitol Hill.*

2. ***No bidding activity:*** *No activity on Auction for "antique chair" item since last update. What do you want to do?*

    a. *No Action, Bid +10$, Bid +20$, Exit Auction*

3. ***New bidder entered auction room:*** *3 new bidders enter auction for "antique chair" item. What do you want to do?*

    a. *No Action, Bid 10$, Bid 20$, Exit Auction*

4. ***Local News: Furniture return of investment increases.*** *Antique chairs gain popularity with interior designers. Antique chairs are now on the top 10 best acquisitions. Source: online auctions.*

A P P E N D I X

C

WEB BROWSING ACTIVITY STUDIES

C.1 WEB SEARCH RESULTS EXPERIMENT INSTRUCTIONS

Please use Google to find the information or webpage for the following scenarios. Tell the experimenter when you have found the results.

NOTES:
Some links in the results page might not work. Just click back and try a different link.
When using the "Back" button, you will be presented with a dialog box, press OK and continoue with the experiment.

SCENARIOS

1. Finding People
You are interested in contacting an old professor to ask for a recommendation letter. Eavan Boland is a poet with whom you worked for some time, but you lost track of her. You should find out what she is up to now and get her email address.

2. Buying products
You are interested in buying tires for your Toyota Corolla 1998 and want to find information about what tires you should buy.

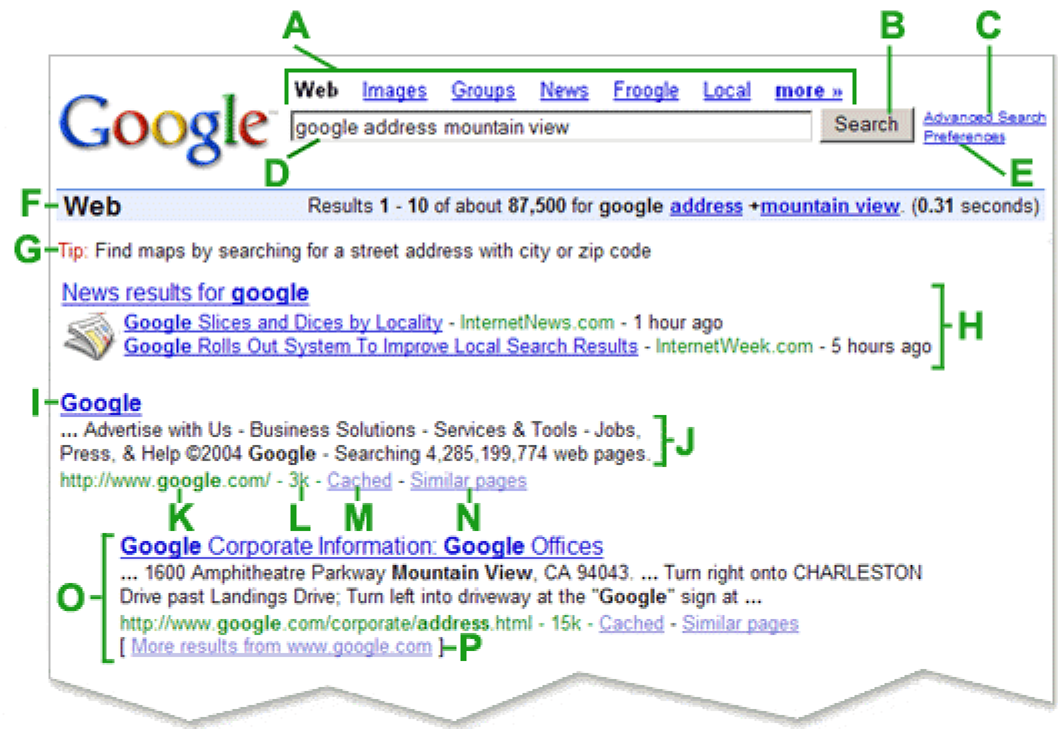
3. You are trying to locate and old classmate from Texas University. The last thing you hear from him was that he was working on weather related projects. Find where is he working now, his work and contact information.
Please use these keywords: **Robert Jacob Texas**

4. Your friend is interested in buying the best digital camera on the market, but at an affordable price and he asked you for your opinion and recommendation. He also asked you to suggest him a website with a good reputation where he could find a Cannon Powershot SD400 camera at a good price.
Please use these keywords: **digital cameras**

5. Perform a free search for anything you want.

6. Perform a free search for anything you want.

C.2 GOOGLE HOME PAGE DESCRIPTION



A. **Top links**
Click the link for the Google service you want to use. You can search the web, look for images, browse Google Groups (Usenet discussion archive), or use Froogle to search for products.

B. **Google search button**
Click on this button to submit another search query. You can also submit your query by hitting the 'Enter' key.

C. **Advanced search**
This links to a page on which you can do more precise searches.

D. **Search field**
To do a search on Google, just type in a few descriptive search terms, then hit "Enter" or click on the "Google Search" button.

E. **Preferences**
This links to a page that lets you set your personal search preferences, including your language, the number of results per page, and whether you want your search results screened by our SafeSearch filter to avoid seeing adult material.

F. **Statistics bar**
This line describes your search and indicates the total number of results, as well as how long the search took to complete.

G. **Tip**
Information that helps you search more efficiently and effectively by pointing out Google features and tools that might improve the query you just made.

H. **OneBox results**
Google's search technology finds many sources of specialized information. Those that are most relevant to your search are included at the top of your search results. Typical onebox results include news, stock quotes, weather and local websites related to your search.

I. **Page title**

The first line of any search result item is the title of the web page we found. If you see a URL instead of a title, then either the page has no title or we haven't yet indexed that page's full content, but its place in our index still tells us that it's a good match for your query.

J. **Text below the title**
This is an excerpt from the result page with your query terms are bolded. If we expanded the range of your search using stemming technology, the variations of your search terms that we searched for will also be bolded.

K. **URL of result**
This is the web address of the returned result.

L. **Size**
This number is the size of the text portion of the web page, and gives you some idea of how quickly it might display. You won't see a size figure for sites that we haven't yet indexed.

M. **Cached**
Clicking this link will show you the contents of the web page when we last indexed it. If for some reason the site link doesn't connect you to the current page, you might still find the information you need on the cached version.

N. **Similar pages**
When you select the Similar Pages link for a particular result, Google automatically scouts the Web for pages that are related to this result.

O. **Indented result**
When Google finds multiple results from the same website, the most relevant result is listed first, with other relevant pages from that site indented below it.

P. **More results**
If we find more than two results from the same site, the remaining results can be accessed by clicking on the "More results from..." link.

http://www.google.com/help/interpret.html

C.3 WEB BROWSING TRACKING FORM

Subject # ___
Free Search #__ Prompted Search #__ Simulated Search #__ CLUSTER ID_____
Search:____
Activity:____
____
____
Clicked link:____
Comments:____

Time	Actions								
:05	Ty	R(w/)	R(w/o)	T(m)	T(s)	Sc	Cl	Ba	Wa
:10	Ty	R(w/)	R(w/o)	T(m)	T(s)	Sc	Cl	Ba	Wa
:20	Ty	R(w/)	R(w/o)	T(m)	T(s)	Sc	Cl	Ba	Wa
:30	Ty	R(w/)	R(w/o)	T(m)	T(s)	Sc	Cl	Ba	Wa
:40	Ty	R(w/)	R(w/o)	T(m)	T(s)	Sc	Cl	Ba	Wa
:50	Ty	R(w/)	R(w/o)	T(m)	T(s)	Sc	Cl	Ba	Wa
1:00	Ty	R(w/)	R(w/o)	T(m)	T(s)	Sc	Cl	Ba	Wa

Google Results Included:

(Search 1)	Advertisement	Google Box	Sponsored links	Suggested Links
(Search 2)	Advertisement	Google Box	Sponsored links	Suggested Links

C.4 USER BEHAVIOR LABELING GUIDELINES

R= Reading (Visual Mostly)
There is no mouse indication. You just know that the user is reading because of the way they move their eyes, or what they tell you.

Z = Pause (or thinking)
Sometimes people take some time to think, they normally park their moue while doing this. They might say something like mmmmm??. Some other people move their mouse randomly while thinking.

I = re-reading instructions on paper
Sometimes people forget the instructions, and they stop to re-read the instructions on paper. It is very important to make sure they read the instructions before diving into the task

M= Mouse move
This is a key behavior. People use the mouse as an aid to read, or moving it while reading, following the text laterally, or moving the mouse slowly vertically.

S Scrolling
This is very straight forward. However, since we are interested in high level behaviors, M, R, take precedence over this one. SO if you see someone scrolling and reading at the same time, make sure you label the activity as reading. You could also click reading-scrolling-reading-scrolling multiple times. This is tricky, but the most useful info.

K = Key usage
Some people use the arrows, PgDown buttons to navigate.

F = ctrl+ F to find text
People should be instructed not to use Ctrl+F since the task is about reading the webpage. SO this shouldn’t happen.

A P P E N D I X

D

DISRUPTION EXPERIMENTS

D.1 EXPERIMENT INSTRUCTIONS

These instructions were be available as reference through the duration of the experiment.

Scenario

You work for a large supply company in the purchasing department. The company is testing a system that assigns different type of customers to different employees through the day, as to maintain a balanced workload. The company has also implemented an Instant Messaging system that allows its employees to share pricing information with one another. Sharing information benefits the company and you might receive a bonus based on the company performance.

The new system classifies and sorts customer emails depending on the type of service requested and the customer's demands. The system places emails in separate folders:

- **HIGH ACCURACY** for customers demanding high accuracy levels, up-to-date prices, *error-free* orders and a high quality of service.
- **URGENT** for customers demanding their orders to be processed as fast as possible. Timely processing is their number one priority.
- **MED-VOLUME** for returning customers from medium-big companies interested in meeting purchasing quotas. They want all
- **LOW PRIORITY** for low volume customers with no accuracy, time or quota restrictions. These customers have no specific demands.

Task Details (Version 4)

As part of your job, you receive many requests from multiple customers over email regarding products they would like to buy and get a price quote (each customer has a Word file—**William Smith.doc**). Your task is to process each customer file with type of service necessary based on folder they were assigned to (customer type).

Each of these folders and each email must be completed before moving to the next one.

Your task is to scan the email and become familiar with the items in the order. Then open the customer file available under "**My Documents**" in the corresponding folder (**High Accuracy, Urgent, Med-Volume, Low Priority)**. Find the online catalog price for each of the items requested by the customer and update the Word document. The word file should include price, quantity, subtotals per product, grand total, and the difference from available budge (see *William Smith.doc* template on next page). Feel free to use the calculator software provided to perform calculations. Reply to the customer's email with a brief explanation about any changes to the order, and a list of unavailable items if there are any. Sign and send the email (see email template in next page)

Your job also includes making decisions so that customers are able to buy as many items as possible while accommodating their preferred products; all within their requirements (quality, quantity, etc). If the budget is exceeded, you should suggest some items in the email that could be removed from the order (use approximate values). Customers rely on your information and intuition and will ultimately place an order based on your suggestions, however, it is in the best interest of the company to keep your customers satisfied. Customers with their demands met improve the company's economic performance. *Please keep this in mind when completing your tasks.*

Instant Messaging System

The Instant Messaging system allows employees to share pricing information with one another. As you work, the IM system provides price updates that may have changed from product catalogs. Instant Messages are designed to save you time by providing you with the most updated information. Use these updates updates whenever possible. Company policy specifies that quotes should be updated if the change exceeds **$10** or if the customer requested accurate quotes. However no updates are necessary for previously processed emails/orders.

The IM system also requests pricing information from you, which is important to the success of the task. Responding to IM messages is part of you job description; however, you should manage Instant Messages in order to best satisfy your customers' requirements.

IM messages are accompanied with a pop-up notification window that reads "*You have a New IM message*". If you want to access the IM message, clicking on the IM notification will bring up the IM message details and open a chat window. Clicking on the notification also indicates [the IM system] that you were interested in the IM and that you intent to follow-up on the message content. If the is not related to your task, or if you don't have time to act on it, you can close the chat window to disregard the message.

IM notifications will automatically fade-away after a few seconds if not attended to as not to distract you. If you need to access previous messages, then you can open a message history file available in 'My Documents'.

Here are some message examples:
"*Please provide a price update for AAA product*" ------ Reply if you have the information requested at hand, or if you have time to respond while processing your current customer based on his preferences.
"*Price Update for AAA product. Now $15.99*" -------- Update the word document to reflect price changes depending on your current customer preferences.

William Smith.doc Template Example

Customer	Category	Product	Qty	Price	Subtotal
William Smith	Books				
		Harry Potter Boxset(Books 1-5)	1	24.5	24.5
		The Da Vinci Code	1	5.5	*5.5*
					30
	Backpacks				
		OGIO Friction	1	50	*50*
		JanSport Interface	1	40	*40*
					90
				TOTAL	**120**
				BUDGET	**130**
				Total - Budget	*-10*

Customer Response Email Template

```
Date: Tue, 22 Aug 2006 01:41:18 -0400
From: Name <XXXX@media.mit.edu>
To: Customer Ordering Services <arroyoernesto@hotmail.com>
Subject: Customer Response Template Email

Dear William Smith,

Your order was processed successfully.
[Rationale for changes to the order here]

All items requested were added to your purchase order.

Best regards,

Ordering Processing Staff

--------------06090406010305
```

Practice

A practice run will allow you to familiarize with the applications and catalogs used for the task. The email client has several e-mail orders folders. Scan or read the email inside the **00_Practice** folder and familiarize yourself with the items being ordered. Process orders in the email folder on a first-come-first-serve basis one at a time and according to the type of customers in that folder.

Take as much time as needed, feel free to review these instructions and ask the experimenter any questions that might arise at any time.

D.2 NASA-TLX WORKLOAD QUESTIONNAIRE

1. Introduction

NASA TLX is a method used to determine subjective workload ratings for a given task.

2. Instructions

After each experimental session we will measure the "workload" you experienced during the session. Workload is difficult to define but can be seen as made up of different factors (e.g. physical or mental components). A set of six rating scales has been developed to evaluate the workload experienced during different tasks. Please read the descriptions of the scales carefully. If you have a question about any of the scales, please ask me about it. It is extremely important that they be clear to you. You will again be given the descriptions to refer to at the end of each experimental trial. .

3. NASA TLX Rating Scale Definitions

RATING SCALE DEFINITIONS

Title	**Endpoints**	**Descriptions**
MENTAL DEMAND	*Low/High*	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
DISRUPTION	*Low/High*	How disruptive to your task were the IM messages presented during the experiment?
TEMPORAL DEMAND	*Low/High*	How much time pressure did you feel due to the task elements? Was your pace slow and leisurely or rapid and frantic?
EFFORT	*Low/High*	How hard did you have to work mentally to accomplish your level of performance?
PERFORMANCE	*good/poor*	How successful do you think you were in accomplishing the goals of the task set by the company (or yourself)? How satisfied were you with your performance in accomplishing these goals?
FRUSTRATION LEVEL	*Low/High*	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

P

4. NASA TLX Participant Rating Form

After each experimental trial, the six rating scales will be presented to you. You will be asked to rate your feelings on each of them by marking a point on the scale.

Mental Demand

Low High

Disruption

Low High

Temporal Demand

Low High

Effort

Low High

Performance

Low High

Frustration

Low High

D.3 USER EXPERIENCE QUESTIONNAIRE

Full Name:		
Age:	e-mail:	
	Gender: Male / Female	Date: Time:
	Notes:	

Questionnaire (Part 1)	Rating				
	Never	1 – 2 times	3 – 10 times	11 -50 times	51 – 100 times
HOW OFTEN HAVE YOU DONE EACH OF THE FOLLOWING					
Written a document using a word processing program (e.g., MS Word, WordPerfect)					
Played a game on a personal compute (Other than Solitaire and hearts)					
Organized information using a database (e.g., Access, Oracle, SQL)					
Written a computer program (e.g., using Scheme, Java, C, PERL, etc)					
Used a computerized spreadsheet (e.g. MS Excel)					
Used a computer assisted design program (e.g, Autocad, Omax)					
Used a statistical program to analyze data					
Used the internet to search for information					

Questionnaire (Part 2)	Rating				
	Novice	Beginner	Intermediate	Expert	Guru
Please rate your confidence in the ability to use the following systems					
Software development					
Web browser					
Word processor					
Graphics/drawing program					
Electronic mail					
Spreadsheet					
Database					

APPENDIX

E

DISRUPTION MANAGER

E.1 SAMPLE MANAGER'S ACTIONS

Email Number	**Interruption Number**	**Related or Unrelated**	**Ongoing Activity**	**Action**
1	1	R	Looking for price	Presented after small changes in activity
1	2	U	Any	Delayed until email sent
1	3	U	Any	Delayed until email sent
2	1	U	Any	Delayed until task switch
2	2	U	Any	Delayed until question answered
2	3	R		Presented after small changes in activity
3	2	R		Presented after small changes in activity
4	1	R		Presented after small changes in activity
4	2	U		Delayed until question answered

E.2 INSTANT MESSAGE RELEVANCE TO EMAIL AND WEBPAGE

IM_QUALITY		**R**	**RE**	**IE**	**RW**	**IW**
Please provide price for "Sports Body Ball" from The Sports Authority Direct Catalog	1	Y	10.19		4.006	0
Price Update for "Vans Stroke Tee". Now $17.99.	1	N		0	0	0
Please provide price for "brilliance side table" from the Chiasso Catalog.	1	N		0.918	0	0
Price Update for "Steam Wizard Sanitizing Steamer". Now $115.78	2	Y	9.028		4.167	0
Please provide price for "hi-lo cocktail table" from the Chiasso Catalog.	2	N		2.525	0	0
Price Update for "DVS Metal Tee". Now $19.99.	2	N		0	0	0
Please provide price for "Loft Three-Shelf Cart" from the Crate and Barrel Catalog	3	Y	14.64		7.039	0
Price Update for "spring tealight holders". Now $28.	3	N		0	0	0
Please provide price for "hanging ball clock" from the Chiasso Catalog.	3	N		3.044	0	0
Price Update for "HydroSilk Women's Long-Sleeve". Now $51.34.	4	Y	8.594		3.827	0
Please provide price for "Etnies Icon Tee" from the Eastbay Action Clothing Catalog.	4	N		2.885	0	0
Price Update for "concentrics rug". Now $298.	4	N		0	0	0
			10.61	1.171 5	4.759 75	0
IM_URGENT						
Price Update for "Napa Wine Table". Now $131.04	1	Y	8.5		10.88	0
Please provide price for "Adidas Team Fleece Hood" from the Eastbay Action Clothing Catalog.	1	N		0.742	0	0
Price Update for "tall tribal sculpture". Now $98	1	N		0	0	0
Price Update for "Magnetic DartBoard". Now $20.52	2	Y	7.14		9.247	0
Please provide price for "arise table" from the Chiasso Catalog.	2	N		1.061	0	0
Price Update for "seashore throw pillows". Now $68.	2	N		0	0	0
Please provide price for "North Portable Folding Picnic Table" from the Sports Authority Direct Catalog	3	Y	12.5		5.303	0
Price Update for "Dragon Logo Beanie Hat". Now $17.99.	3	N		0	0.673	0.864
Please provide price for "zhenni floor lamp" from the Chiasso Catalog.	3	N		2.178	0	0
Please provide price for "100 Mini Cell Phone Case" from the NRS Water Sports Equipment Catalog	4	Y	17.71		8.727	0
Price Update for "gridline short bookcase". Now $398.	4	N		0	0	0
Please provide price for "fusion rug" from the Chiasso Catalog.	4	N		3.511	0	0
			11.46	0.936	8.539	0.072

Relevance
RE = Score for Relevant Email
IE = Score for Irrelevant Email
EW = Score for Relevant Webpage
IW = Score for Irrelevant Webpage