A comparison of screen size and interaction technique: Examining execution times on the smartphone, tablet and traditional desktop computer

A Literature Review

University of North Carolina at Chapel Hill
School of Library and Information Science

Laura Haak Marcial
28 September 2010

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEFINITIONS

Concept	Definition
cellphone	cellular telephone, also known as a mobile phone and differentiated somewhat from smartphones and PDAs
cellular/mobile networks	radio network distributed over land using base stations/cell towers to create an area of radio availability for portable transceivers (like mobile devices) Generations of this technology include: 1G (analog, 1981) > 2G (digital, 1992) > 3G (multi-media support and 200kb/s, 2002) > 4G (IP packet switched, gigabit speeds, multi-carrier, to be released))
clamshell	mobile phone form factor that closes via a hinge (flip)
CMN GOMS	original Card Moran and Newell Goals Operators Methods and Selection Rules Theory; used to disambiguate other versions of GOMS
CogTool	CogTool is a general purpose UI prototyping tool with a difference - it automatically evaluates your design with a predictive human performance model (a "cognitive crash dummy"). (http://cogtool.hcii.cs.cmu.edu/)
CPM GOMS	Bonnie John and John Kieras's proposed revision to GOMS to include Cognitive-Perceptual Motor attributes
DigitalDesk	developed by Pierre Wellner in 1991, the DigitalDesk included a typical work desk, a camera, an LED pen and (in a more advanced version) a computer-driven projector
display	synonymous with monitor, screen, CRT (cathode ray tube), etc.
gaming device	a mobile device oriented toward gaming
goal	according to Xie (2009), goal can be any of the following: a. Long-term goal refers to a user's personal goal that they will pursue for a long time, such as professional achievement (e.g. doctorate degree). b. Leading search goal refers to a user's current task-related goal that leads to a search (e.g. writing a paper). c. Current search goal refers to the specific search results a user intends to obtain (e.g. find relevant literature on task). d. Interactive intentions refer to sub-goals that a user must achieve to accomplish their current search goal. Task and goal are inseparable in the information-seeking and -retrieving process.
GOMS	Goals Operators Methods Selection Rules
identical query	a query within a session that is a copy of a previous query within that session Jansen, Spink and Pedersen (2005) based on Jansen and Pooch (2001)
information behavior	the totality of human behavior in relation to sources and channels of information, including both active and passive information seeking, and information use. Thus, it includes face- to-face communication with others, as well as the passive reception of information as in, for example, watching TV advertisements, without any intention to act on the information given. Wilson (2000)
information searching behavior	the 'micro-level' of behavior employed by the searcher in interacting with information systems of all kinds. It consists of all the interactions with the system, whether at the level of human computer interaction (for example, use of the mouse and clicks on links) or at the intellectual level (for example, adopting a Boolean search strategy or determining the criteria for deciding which of two books selected from adjacent places on a library shelf is most useful), which will also involve mental acts, such as judging the relevance of data or information retrieved. Wilson (2000)
information seeking behavior	the purposive seeking for information as a consequence of a need to satisfy some goal. In the course of seeking, the individual may interact with manual information systems (such as a newspaper or a library), or with computer-based systems (such as the World Wide Web). Wilson (2000)
information use behavior	consists of the physical and mental acts involved in incorporating the information found into the person's existing knowledge base. It may involve, therefore, physical acts such as marking sections in a text to note their importance or significance, as well as mental acts that involve, for example,

	comparison of new information with existing knowledge. Wilson (2000)
initial query	the first query submitted in a session Jansen, Spink and Pedersen (2005) based on Jansen and Pooch (2001)
initial query stage	stage in which the search strategy is constructed Rieh and Xie (2006) based on Efthimiadia (1993)
search intention	the "goal" behind search intent; not to be confused with user intent or query suggestion
interaction device	device used to interact in a human computer system; includes input and output devices
interaction style	according to Shneiderman (1997) and Preece (1994), the basic forms of interaction style include command language, natural language, form fill in, menu selection, direct manipulation, and virtual reality
interaction technique	method of using an interaction device to perform a task
ISO 9241-11	usability standard which provides guidance on the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.
keyboard (varieties)	main input device for computers; based on a typewriter keyboard, a set of buttons or keys which produce or correspond with letters, numbers, symbols or actions when pushed or touched. Variants include: full sized, laptop sized, thumb sized, numeric, chorded, soft, and projection among others.
keyboard (layouts)	any specific mechanical (ANSI, ISO, JIS), visual or functional layout of keys. Typically, the mechanical layouts are very similar across devices and languages with the exception of mobile devices which have several variants. Visual layout is language dependent and functional layout is software dependent. For most Latin script languages a QWERTY style keyboard is used.
keypad (numeric and alphanumeric)	aet of buttons arranged in a block or pad with number letters and some symbols. Modeled after telephones, use of keypads on a mobile phone for text input requires either single tap, multi-tap or predictive text entry.
KLM	Keystroke Level Model, the simplest of the GOMS variants
Microsoft Surface	Microsoft's multi-touch enabled computing surface released in 2008 and designed for multi-user gestural recognition computing.
mobile device	any of a number of devices designed for use in a mobile context; can range from
multimedia device	mobile device which supports interaction with multimedia (music, images, movies, and games, etc.), typically an iPod or similar
multi-point	interface which allows user to interact via multiple points—allows parallel processing of information from multiple points and supports bimanual input; see also multi-touch
multi-touch	describes a touchscreen capable of receiving input from three or more distinct touches; has properties of multi-point devices
netbook	small lightweight and inexpensive laptop designed for web based use (to augment other computing devices)
NGOMSL	Natural Language Goals Operators Methods and Selection Rules
PDA	personal digital assistant, the precursor to modern smartphones
predictive text	single keypress of the keypad of a mobile phone for commonly used words (rather than multipress)
QGOMS	Quick (and dirty) Goals Operators Methods and Selection Rules
query	the entire string of terms submitted by a searcher in a given instance Jansen, Spink and Pedersen (2005) based on Jansen and Pooch (2001)
reformulation query stage	stage in which the initial query is adjusted manually or with the assistance of a system Rieh and Xie (2006) based on Efthimiadia (1993)
repeat query	a query submitted more than once, irrespective of the user Jansen, Spink and Pedersen (2005) based on Jansen and Pooch (2001)
search move	an identifiable thought or action that is a part of information searching Bates (1990)
search stratagem	a complex of a number of moves and/or tactics and generally involves both a particular identified information search domain anticipated to be productive by the searcher, and a mode of tackling the particular file organization of that domain Bates (1990)

search strategy	a plan which may contain moves, tactics and/or strategems for an entire information search Bates (1990)
search tactic	a set of search moves that are temporally and semantically related Bates (1990)
search task	a task that determines what a user is searching for (Xie, 2009)
search term	specific word or phrase used in a search
session	the entire series of queries submitted by a user during one interaction with the Web search engine Jansen, Spink and Pedersen (2005) based on Jansen and Pooch (2001)
simple search	can mean either an uncomplicated search goal or an uncomplicated search term or an uncomplicated search process (Jansen, Booth and Smith, 2009)
smartphone	synonymous with converged device, preceded by a PDA and differentiated from a cellphone
softkey	programmable key such as the F key of a keyboard; typically in use near the display of a mobile device
tablet computer	a laptop equipped with a stylus and/or touchscreen
task	what someone does to achieve a goal (Xie, 2009 referencing Hackos and Redish, 1998, p. 56)
term	any series of characters separated by white space or other separator Jansen, Spink and Pedersen (2005) based on Jansen and Pooch (2001)
touch sensitive	responding to touch as in touchscreens
transaction log analysis	the study of electronically recorded interactions between on-line information retrieval systems and the persons who search for information found in those systems. Jansen, Taksa and Spink (2009) based on Peters (1993)
word completion	automatic completion of commonly used words in text entry (or suggestion of possible terms in a list)
word prediction	automatic prediction of possible words from a list that refines as the user types
work task	a work task represents a task that leads to information searching (Xie, 2009)
wifi	term used in advertising any wireless local area network capable device based on the IEEE 802.11 standard

1. INTRODUCTION

An information age is upon us. In no other sphere is this as evident as in the world of mobile devices. From the quantity of available hardware, software and network options to the magnitude of data being generated by mobile devices right now across the world, nothing else compares. The notion of a personal communication device has universal appeal, regardless of an individual's level of prior experience, income or education. The mobile device has reached people in places where technology has not gone before from the African savannah to the mountains of Nepal. Designers, manufacturers and researchers alike proclaim ease of use, user centered design, focus on the user experience and technological improvements in battery life, display resolution, and wireless network infrastructure to be key factors in the uptake of mobile devices. They are indeed, the first piece of technology of any kind to break through to the 'bottom billion' so to speak. Will mobile devices completely supplant traditional computing devices and transform our current notions of how computers should look, feel and be interacted with?

Problem: The Technology Paradox

In their 1993 article, "An Agenda for Human-Computer Interaction Research: Interaction Styles and Input/Output Devices", Jacob, Leggett, Myers and Pausch state that "The bottleneck in improving the usefulness of interactive systems increasingly lies not in performing the processing task itself but in communicating requests and results between the system and its user". This statement presaged the current dilemma in mobile computing today.

According to Jacob et al. (1993), we continue to struggle with the 'demand-pull' versus 'technology-push' of user interface design in the development of new and mobile computing technologies. Driven by market forces and perceived user preferences, not necessarily human-centric design, technology developers are almost ambivalently barreling down two paths: one where personal technology devices appear to be converging into a single primary interface versus one where they are diverging and becoming increasingly specialized, even personalized. More and more designers are suffering from *featuritis* (Chang et al., 2007) and users from feature fatigue (Thompson, Hamilton and Rust, 2005). From skins and ringtones to Global Positioning Systems (GPS) and altimeters, the sea of possible ways to use and customize personal computing devices is limitless. Though the mouse and keyboard have been around as the primary interaction devices for desktop and laptop systems, these systems and their interaction devices have been eclipsed by the advent of small form factor computing and, as a consequence of size, a wide variety of new and unproven interaction styles. As complexity increases, the synergy between devices breaks down. For the typical computer user this means that she can no longer take advantage of the skills she has developed to interact with one system when interacting with another (Yamashita, Barendregt, and Fjeld 2007). This also works in reverse; a newer device with improved interaction may precipitate frustration when integrated into a user's computing suite because other devices in the network do not perform to the same standard. Because a significant number of users now own and [want to] synergistically operate between a desktop, laptop, tablet, and sometimes multiple mobile devices, identifying a **simple straightforward set of interface standards** could significantly enhance this interaction.

With the advent of the personal computer came the keyboard, mouse and display. Since that time, the variations on these three forms of input have been substantial. From the joystick to the Wii glove, modern technology has sought to translate human gestures into recognizable and meaningful human-computer interactions. Though the focus to date has been on keyboards, typing, mousing and a visual interface, more recent trends are focusing on handwriting and voice recognition, gesturing and multi-touch interaction as well as virtual reality and projection systems. Indeed, there is a need for a '**paradigm shift**' in interaction styles, techniques and devices where mobility, ubiquity and computing devices are concerned (Lumsden and Brewster, 2003).

Beyond mere adoption, technological advancements have pushed us toward an ever-increasing paradox: the challenge of complexity. Mahler and Weber's (2008) 'Paradox of Technology', (see Figure 1) illustrates how we are bound by the continual development of new features to solve old problems. Mobile devices have become increasingly complex in an attempt to address the tension between small display size and the resulting interaction style for a mobile context.

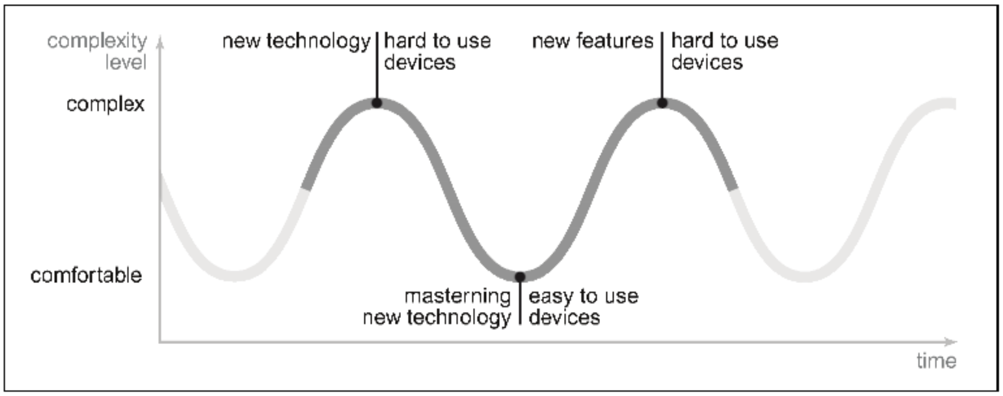

Figure 1. Titled "The Paradox of Technology" from Mahler and Weber, 2008 citing Norman, 1988

The TabletPC, offering a handwriting recognition feature attractive to professionals, made a significant contribution to the long-standing interaction barrier the keyboard posed, particularly for drawing interactions. Research involving the use of TabletPC devices has been done in medicine, among the military and more recently, in teaching and education. Though certain situations appeared to be more conducive to the use of handwriting recognition, problems with quality and speed have limited uptake. Moreover, with the advent of the keyboard, most users who become proficient with it stray away from writing by hand and increasingly use the computer keyboard as their primary writing tool. It may even be a serious consideration that the issue of handwriting recognition will dissolve as a generationally-dependent problem. As access speeds and processing speeds dwindle, the perception that interaction should be even more instantaneous increases. This is causing a general shift toward advances in voice recognition technology and smarter gestural interaction.

It is broadly recognized that the greatest weakness of mobile devices is directly related to their greatest strength: small display size. **Despite their unprecedented uptake and use, it remains unclear whether mobile devices are really capable of supplanting traditional laptop or desktop computers for a significant number of tasks.** While most manufacturers and wireless carriers have overcome the limitations of battery life, operating system failures, software availability, network availability (data and voice) and cost, significant issues still remain. Increasingly, consumers are experiencing feature fatigue—frustration with the complexity that additional features can promote, interaction issues (like that of a virtual keyboard) and display size limitations. There remain a significant number of tasks which seem quite difficult to perform on a mobile device: composing music or a term paper, working on architectural drawings, performing a financial analysis, reading an X-ray, conducting scholarly research, etc. These, even, are extreme examples. Over the last decade researchers have been trying to better understand use of mobile devices for internet searching. In so doing, it is clear that both display size and interaction issues remain the single biggest barriers to extension of use beyond simpler personal information management tasks such as email, texting, maintaining contact lists and a calendar. Still, mobile devices are poised to become a primary means of accessing the internet.

Modeled after the desktop calculator, music players like the Sony Walkman introduced in 1979-80 made the music experience highly personal and private. The advent of personal organizers like the Apple Newton (way ahead of its time in 1993) and the Palm Pilot (1996), took the notion of a pocket –sized paper-based organizer and calendar and made it digital. Initially, the one big drawback to this was the lack of a connection to any network and some technical difficulty synchronizing data with the desktop. Moreover, once the PDA existed, it was evident that the newly popularized cellular phone could easily be merged with a PDA to create a 'converged device'. Though the first of these emerged in 1992 (IBM Simon), their success in US markets came with the introduction of the Handspring Treo in 2002 which merged the popular Palm OS features, phone features and a full keyboard with wireless web browsing.

Just prior to the launch of the Handspring Treo was that of the first Apple iPod device in 2001. These thumbwheel driven music devices quickly morphed into full fledged multimedia devices facilitating listening to music, watching videos, and managing related collections. The 'personal' nature of these devices and their novel interaction style made them extremely popular and set a new precedent for ease of use. Soon after came the launch of a converged smartphone and multimedia player with a new fundamental interaction style—multi-touch. Enter the iPhone (and iPod Touch), the first device to attempt to blend personal multimedia capabilities with a sophisticated mobile phone and all of the features of a PDA, in a sleek package with a mostly smooth and intuitive interaction style.

With the advent of multi-touch devices with small form factor, portability and high resolution displays like that of the iPhone/iPod touch, a question about computing device replacement or surrogacy has arisen. The reason this technology is pacesetting is not due to increased screen real estate (which remains small), rather it is the pinch and zoom resizing options that make web browsing with or without user interface modifications finally plausible. Since the first appearance of web browsing in the mobile environment, efforts have been made to improve the user experience through design of web pages, software, and interaction devices. Today, the topic has shifted to fundamental improvements in device design and human computer interactions which would facilitate improved interaction without requiring the tailoring of the content for different display devices.

For the last decade, mobile phone technologies have been the fastest growing segment of the technology market. While the debate about whether computing technologies are converging into a single device for the majority of users or diverging into increasingly specialized and sophisticated tools, the issue of adoption remains centered on two pivotal human computer interaction factors: display size and interaction style. For some, the availability of features in any given computing device today can be so overwhelming as to cause feature fatigue. This combined with myriad differences in display size and interaction styles creates an environment where research and development are consistently confounded by significant variability among devices within these factors alone. **The research outlined here seeks to understand more about the execution portion of task performance on a range of computing devices**.

Advantages and Disadvantages

It may be hard to overstate the advantages mobile devices have afforded the typical individual since they first became available. Initially used primarily for emergency purposes, use has grown to a point where an increasing number of households maintain only a mobile phone for household use and no longer have a landline at all. Remote areas of the globe where infrastructure for water/sewer and transportation alone have not been built out much less for telephone and electricity can utilize mobile devices with longer battery life and cellular towers and practically leap beyond these infrastructure hurdles. If information access for all is the goal, mobile devices have done more toward achieving this than any other technology to date.

While the majority of mobile device users may be convinced about their virtue and remain steadfastly optimistic about the improvements they represent, there is a growing concern about their addictive nature. Instant access to information and communications options also means an increase in distraction and perhaps an inability to focus (Wobbrock, 2006; Holleis, Otto, Hussmann, and Schmidt, 2007; Chittaro, 2006; Roto, 2006; Arter, Buchanan and Jones, 2007; Chittaro, 2004). In addition, these 'wearable' devices may also pose an electromagnetic radiation exposure risk in those who use them for long periods of time (Moulder, Foster, Erdreich, and McNamee 2005). If indeed these devices and the ubiquitous electronic access (Mahler and Weber, 2008) they represent is here to stay, these problems and more will need to be addressed.

Mobile Context

"Mobile cellular has been the most rapidly adopted technology in history. Today it is the most popular and widespread personal technology on the planet, with an estimated 4.6 billion subscriptions globally by the end of 2009." (International Telecommunications Union (ITU) Measuring the Information Society (MIS) Executive Summary, 2010, pg. 1) In this information age, the typical consumer is regularly overwhelmed with options for hardware, software and peripherals. The number of available types of mobile devices alone has climbed into the thousands in the US with wide ranges in capability, features, interaction device and style, display size, communications options and form factor. Why then, don't these devices adhere to some basic standards for operation, use and evaluation?

In 2009, a report from the Interactive Data Corporation indicated that the number of people accessing the internet by 2013 will reach 2.2 billion compared with 1.6 billion (24% of Earth's total population at the time) in 2009. (IDC Press Release "Number of Mobile Devices Accessing the Internet Expected to Surpass One Billion by 2013, According to IDC", 2009) The mode of access, traditionally through desktop or laptop computers, however, is changing. According to the same report, "over 450 million users sought access to the Internet through mobile devices this year", the article also asserts that 1 billion mobile devices will be used to access the internet by 2013. (IDC, 2009)

By mid 2009, 90% of US consumers were wireless customers (CTIA—The Wireless Association "CTIA's Semi-Annual Wireless Industry Survey", 2009) and the top selling mobile phone in the US was the Apple iPhone (4%) (NielsenWire, 2009). According to a recent Pew Internet and American Life survey (2009), of the 83% of US adults with cellphones, 35% have accessed the internet via their phone. The same survey indicated that 32% of Americans have used a mobile phone to access the internet and nearly one fifth of Americans use the internet on a mobile device on a typical day. In addition to internet use, the mobile phone is also eclipsing the number of landlines in use in some areas, particularly those where land line infrastructure is poor or nonexistent (ITU MIS Executive Summary, 2010). In the US, a recent CDC study by Blumberg and Luke (2008) indicated that more than one in five American homes had only a wireless phone in 2008. The trend towards reliance on wireless is particularly evident among 18-24 year-olds where one in three lived among the wireless only households. (Blumberg and Luke, 2008)

The first of the cellular networks emerged in the late 1970s and have continued to rapidly evolve from analog (1G) to digital (2G) to wideband mobile (3G) and, recently, broadband mobile (4G) has emerged. The current expectation is that 4G will offer 'anytime, anywhere' access for voice, data and multimedia. Carriers are expected to launch 4G networks in 2011. The figure below from Sharma (2009) helps demonstrate how wireline and mobile technologies have developed and may be developing.

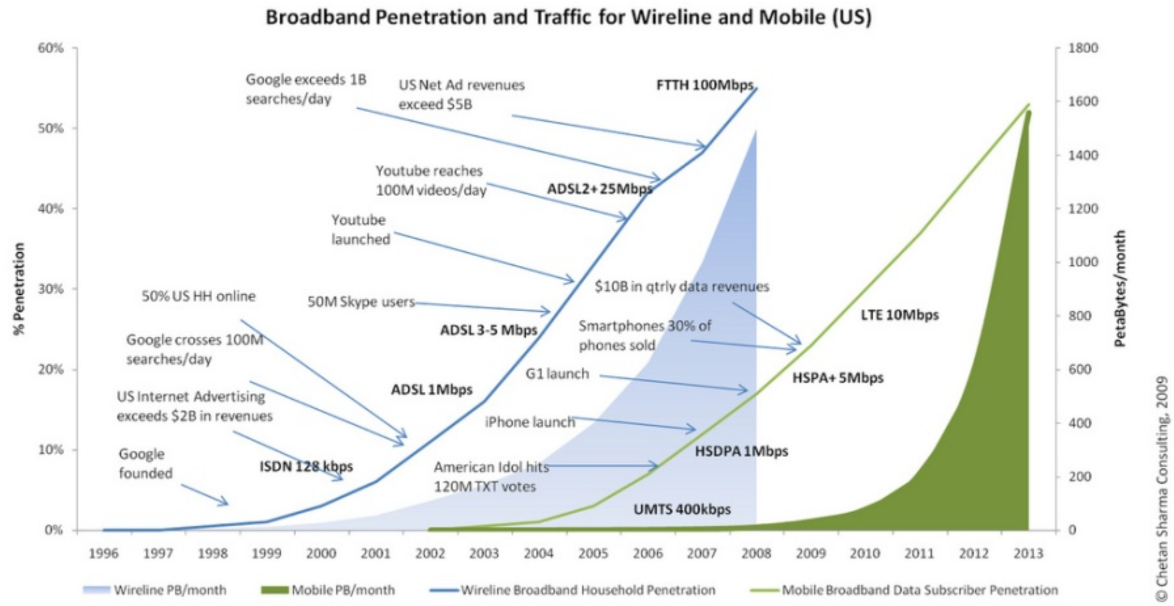

Figure 8. Broadband penetration and traffic for Wireline and Mobile data networks in the US (1996-2013)[29]

[29] Source: Chetan Sharma Consulting. Data Sources: Wireline Traffic Data from Minnesota Internet Traffic Studies (MINTS) - http://www.dtc.umn.edu/mints/igrowth.html. We took the mean of the yearly ranges. Wireline Broadband HouseHold penetration data from Pew Internet - http://www.pewinternet.org. Mobile Data traffic – Chetan Sharma Consulting analysis. Mobile Broadband Data Subscriber Penetration – Chetan Sharma Consulting analysis

Figure 2. From Sharma, 2009

The first mobile phone was used to place a call in 1973 and since the late 1990s, they have been in widespread use even reaching (and exploding in) the bottom of the economic pyramid penetrating markets in sub-Saharan Africa and India starting in 2004. Western Europe (e.g. Nokia in Finland) continues to pioneer modern cell phone use and design but the largest growth in use of mobile devices has occurred in Asia and Africa where growth rates are exponential. Many low resource countries are literally skipping landline phone technology with the development of mobile phone infrastructure bringing both voice communication to places that haven't had ready access as well as internet access, albeit unreliably and without any training or prior exposure.

In the late 1990's mobile phones grew small enough to carry in a pocket. In the early 2000's the PDA or personal digital assistant, a non-networked, non-communication ready device was introduced (though there were iterations of this earlier—Apple Newton). As the decade wore on, the Smartphone was introduced—a converged device which offered both the computing power of a PDA and the communications capability of a mobile phone. The earliest of these were clunky (weight/size), tied to a given operating system (Palm OS) and offered novel (stylus) and often limited styles of interaction (keypad). Today there are thousands of mobile devices available for use ranging from traditional, communications-oriented cell phones to feature phones and smartphones, multimedia players, gaming devices, etc.

A Possible Solution

We are reaching a critical point where the core capabilities of small scale devices has increased (battery life, processing speed, network redundancy and availability and display quality), barriers to adoption have all but vanished and, unlike previously thought, use of these devices appears to be less and less task dependent. Still it remains hard to understand and quantify the ways in which mobile devices perform in a manner commensurate with that of a traditional desktop or laptop system. Many well tested evaluation methods have been used to analyze the usability of desktop computing systems though these same approaches have only just begun to be used in the study of mobile devices.

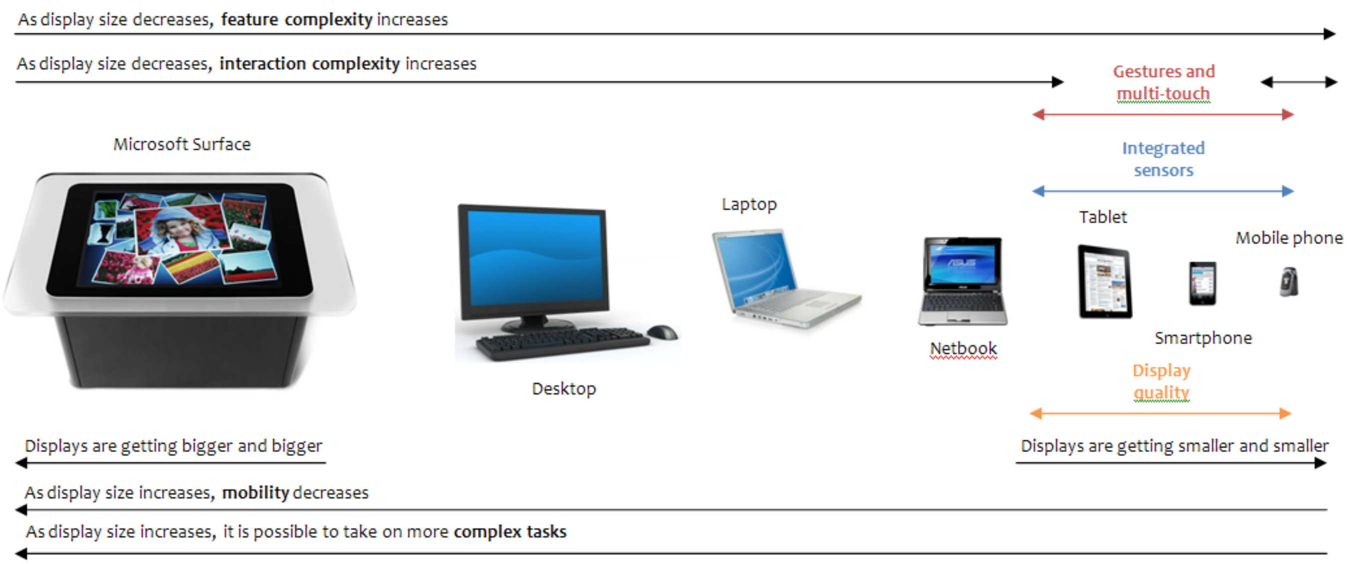

Figure 4. Illustration of the interplay between display size, interaction complexity, mobility and task complexity.

2. DEVICE COMPARISONS

Studies to Date

As a result of trends in the use of technology, in recent years, research on computing devices has evolved toward an increased focus on mobile devices and environments. While early studies focused on issues of adoption and usability, more recently studies have been emerging that attempt to address direct comparisons of specific features between or across a variety of devices.

Many comparison studies look at the advantages and/or disadvantages of using mobile devices versus a technology used prior (often paper and pen), or by comparing a variety of mobile devices to each other and examining issues of adoption or ease of use. Some studies look at specific software comparisons or task comparisons, others at just a single function of the devices (display for example). Fewer studies look carefully at advantages mobile devices may have over other computing methods or at notions of equivalence beyond mobility. This is largely due to the fact that until recently, more factors than display size and interaction style were still central problems with mobile devices. While some issues like battery life and software availability have improved significantly, issues with network availability and processing speed remain, especially for more sophisticated tasks. Context, a crucial element of mobile technologies, is a great challenge to research and evaluation and existing theoretical frameworks are often thought to be inadequate or at least in need of modification(s).

Device variation has remained a significant challenge to software developers and many device comparison studies are aimed at reducing this design burden. Buranatrived and Vickers (2004) examined a similar software application on devices with differing interaction styles and concluded that writing an application once (J2ME) and executing it on different platforms was achievable but that this may inherently be at a cost to usability. Chae and Kim (2003) describe an important relationship between small display size and horizontal depth when designing for small form factor devices. Moreover, their work helps identify significant differences in the perception of users of very small display devices where the cost of navigation is very high versus displays which facilitated fewer than four horizontal depths of navigation. Chan, Lam, Fang, Brzezinski, Zhou, and Xu (2002) compared a wireless application protocol (WAP) browser, a Palm PDA browser and a PocketPC PDA browser in the usability of 10 wireless sites and found that designing for both the novice and experienced user was important, and that flattened sites which emulate that of a desktop system were more easily navigable. Since this early study, many of these issues have been aggressively addressed by software developers.

Domain or task specific comparisons while more replete in the literature suffer from a lack of generalizability. Clegg, Bruciatelli, Domingos, and Jones (2006) conducted an interesting study of digital geological mapping using a Global Positioning System (GPS) on a PDA versus a TabletPC. While PDAs proved convenient for remote mapping, the TabletPC outperformed the PDA in most tasks. Small display and limited processing power were considered to be the significant inhibiting factors. Curran, Woods and O Riordan (2004) conducted a helpful investigation of text input using mobile phones. More importantly, their work drives home a significant point in [at least] mobile technology evolution: that usability often takes a back seat to aesthetics and designers often sacrifice function for forms which seem more appealing. In addition to this, the study highlights the speed versus accuracy tradeoff associated with text input and suggests that text input on devices be tailored to the task at hand and its relevant demand on speed or accuracy. In more recent years, designers and developers have worked hard to understand the task dependency issue with text input and to look at alternative ways of achieving desirable levels of speed and accuracy without regard to task.

Schulz (2007) outlines an important element of study, how well the predictive models of traditional desktop systems apply to the use of mobile devices. In his dissertation work, Schulz created and investigated the use of KLM-Qt a software application designed to facilitate recording of Keystroke Level Model (KLM, see Definitions and Methods sections for more details) operators 'derived from events that are delivered to an interface (Schulz, 2008 p. 4).' The second part of the study used the software to do a comparative study of three different devices, a Greenphone (keypad based) an iPhone and a Neo1973 (both touchscreen based). KLM Qt and hand generated KLM results from a series of 15 routine tasks ranging from creating a new address book entry to adding a meeting date/time were compared across the three devices. It was not possible to run KLM Qt on the iPhone so these models were generated by hand. The Neo1973, like the iPhone employs a touchscreen so a new input operator called "I" was added to the KLM model to handle text composition and 'commitment'. No "multi-touch" gestures were used or studied in this work. Finding suggested that KLM is useful in predicting interaction times on mobile devices though some question remained about the accuracy of the model for all types of interaction types mobile devices typically facilitate.

A recent study by Holleis, et al. (2007) pinpoints an area of particular concern to this work, and that is of the applicability of traditional desktop models of interaction for quality assessment of mobile technologies. While their work finds many parallels, it also identifies areas where improvements to these models could be made to more appropriately describe more sophisticated interactions with mobile technologies. The improvements are outlined as extensions to the GOMS KLM operators and the authors suggest that they apply even to more state of the art mobile devices. Another important point drawn from this research regards the notion of the expert. In this study and others, an expert can be cultivated but it can be more difficult to retrain an expert. This is a significant barrier to making the transition from one device type to another. While many differences exist (by the very nature of mobile devices), the similarities are important and may be suggestive of a trend toward replacement use becoming increasingly plausible. There may also be an underlying 'best in class' set of features or device characteristics which reduce the burden of usability thereby increasing uptake for more sophisticated use(s).

The work of Kamvar, Kellar, Patel and Yu (2009) reinforces the notion that, for higher-end phones, what the user already knows about human computer interaction in terms of personalization and feature set in the desktop system can be leveraged to promote commensurate use of mobile devices. Their work also suggests that no single search interface is appropriate for the range of mobile phones available. The results of this work indicate that iPhone query formation is nearly similar to computer based query formation and that only a small percentage of these searches are locally oriented, that is specific to a geographic area, refuting the notion that mobile searching is largely locally based. Another important finding is that the 'tail', a measure of diversity in query formation, is longer among iPhone users than traditional mobile phone users and is increasingly comparable to computer based users among whom the 'tail' is the longest. The most interesting result in this work suggested that iPhone users, because of improvements in browser capability, connection speeds, display size and resolution and interaction style, typically have more diverse information needs than those of computer based users. All of this is suggestive of the assertion that higher-end mobile devices like the iPhone will increasingly extend their range of applications, and further overlap types of activities, like search tasks, that have traditionally been limited to desktop computer use.

Qiao, Feng and Zhu (2008) take an important approach in surveying existing research into interface design differences between desktop systems and mobile systems with particular emphasis on ways to improve the mobile user interface for web search. In particular, they examine leading display and serial display of query results taking into account optimal display speed and jump length for the human user. They found that providing the ability to pause, continue or stop functions, to allow fast page turning and the option of enlarging what is being displayed are important elements of design for mobile browsing.

Silvey, Macri, Lee and Lobach (2005) conducted a comparison study between a Palm PDA and a Windows TabletPC of the same clinical observation software for eye care. They used both focus groups and usability surveys to determine user preferences with specific regard to care setting. Their findings suggested that the TabletPC was preferred over the PDA and that display size was the most significant limitation of the PDA. Cost and weight were limitations of the TabletPC. While the researchers made every effort to create 'functionally identical' applications for both environments, the environments themselves may be so inherently different that this was itself a major factor. For example, the authors mention that on average a single screen of data on the TabletPC may be broken down into five screens on the PDA.

An interesting new study by Toomey, Ryan, McEntee, Evanoff, Chakraborty, McNulty, Manning, Thomas and Brennan (2010) focuses on a comparison of monitors for emergency radiologic readings of brain CT slices and wrist radiographs. The Dell Axim PDA, the Apple iPod Touch and a secondary-class monitor (first-class being a clinical workstation) were compared. Findings suggest that both the PDA and the iPod Touch performed at least as well as the secondary-class monitor and that the PDA performed better than the monitor on some of the brain images, a statistically significant finding.

Important characteristics

There is a tension between what constitutes real innovation in HCI and what is more aptly described as innovation on a theme. Bill Buxton describes this tension very succinctly in "Surface and Tangible Computing, and the "Small" Matter of People and Design" where he discusses technology innovation as simultaneously being like a rocket and a glacier. The faster work, he says, is more like variations on a theme to reduce cost and the slower technologies, while they may be truly innovative, take so much time to test and prepare for adoption that their 'novelty' has worn off by the time they are readily available.

Indeed, the critical innovations in mobile device design have yet to be made. While the concept of mobile devices is now no longer novel, the variations on a theme that exist in the market today do little to broach the incredible distance between user and device. The comparative studies outlined above draw our attention to the two remaining central issues with increased use of mobile devices: their display size and interaction style. In addition, these comparisons highlight a lack of strong evidence to help guide the use of specific devices in specific settings as well as the selection of the right device for a given task. As developers strive to add features which bring value to the user, devices become increasingly complex in their design and operation. This is often mistaken for improvements in functionality which few users really derive benefit from. On the other end, designers are also keenly aware that users derive significant benefits from efforts made to utilize HCI elements which a broad user base is already familiar with, like the keyboard and mouse.

The introduction of the iPhone and, more recently, the iPad represented two significant shifts in user centered design where mobile devices are concerned. One was the marriage of several key elements of daily human life: communications, personal information management and entertainment, bundled in a way in which users were already somewhat familiar. The other was multi-touch for the masses on a midsize display which was perhaps imperfect but still a significant step in the right direction.

3. DISPLAY SIZE

Since the advent of the personal computer, the display has been a fixed element of human computer interaction. Several key factors have played a role in display development over time: technological advancements (e.g. CRT versus LCD; black and white versus color), cost (miniature versus large scale) and human visual capabilities/limitations (visual acuity & cognition; mobility & distraction; field of view). As the central means of interaction in terms of computing output, displays have been invaluable. With beginnings stemming from a variety of different realms, the display has been refined significantly over time but the basic form and interaction remain largely the same. Moreover, as the central form of computing interaction output, the display remains central despite changes in setting, form factor and mobility in recent years. Though the size, type and feature set (color, etc.) of displays has changed, the mode itself has not. Due largely to changes in the cost of the underlying technology (LCD, CRT, plasma, etc.) and subsequent advancements in size (both larger and smaller), displays run the gamut in size, type and functionality.

One significant development has been the miniaturization of the display. Initially, small displays were very simple like early televisions but with technological improvements, the field has grown unwieldy in terms of display options. Though devices with nearly every possible size of display exist in the computing arena, some small displays are beginning to approximate the quality of more traditional desktop and laptop displays. Display variations for use in cell phones, PDAs and smartphones, have grown most significantly in recent years and soon we can expect to see them employed in a wide variety of wearable devices (watches, etc.). This shift has occurred in large part because of a need to 'go wireless' or become no longer tethered to the desktop.

There are a number of elements which comprise the effectiveness of a given display. The table below outlines key components of evaluating displays and how they are manifest in current practical application:

Table 1. Display Comparison Matrix

Performance Measurement	**Apple 30" Cinema Display** 	**17" Display Dell E178FP** 	**iPad Tablet** 	**Apple iPod Touch** 
Size: typically measured on the diagonal but also in maximum width and height	29.7" 21.3 x 27.2 x 8.46 in	15 x 14.8 x 5.5	9.7" 9.56 x 7.47 x .5 in	3.5" 4.3 x 2.4 x 0.33 in
Type	TFT-LED	LCD display / TFT active matrix	LED-Backlit IPS Display	LCD color transflective TFT display
Support Multi-touch	No	No	Yes	Yes
Aspect ratio: ratio of width/height, typically 4:3	4:3, 16:9	5:4	4:3	2 (horiz) 3 (vert)
Field of view: extent of observable area	55°	*Not indicated*	*Not indicated*	*Not indicated*
Resolution: in pixels or dpi?	2560 x 1600	1280 x 1024 / 75 Hz	1024-by-768-pixel resolution at 132 pixels per inch (ppi)	320 × 480 pixel resolution at 163 pixels per inch (ppi)
Dot pitch or pixel pitch ratio: distance between pixels of the same color, the smaller the better	.250 mm	.264mm	132 pixels/inch	164.6 pixels/inch (0.15428 millimetre/pixel)
Color range	16.7 million	24-bit (16.7 million colors)	*Not indicated*	262,144-color
Refresh rate: # times in a	60 Hz	Max Sync Rate (V x H):	60 Hz	2.5 Mbps, 30 frames

second that a display is illuminated (max by response time)		76 Hz x 81 kHz		per second
Response time: time for a pixel to go from black (active) to white (inactive)	16 Ms	5 Ms	*Not indicated*	*Not indicated*
Contrast ratio: luminosity of brightest color (white) to darkest color (black)	700:1	800:1	*Not indicated*	*Not indicated*
Luminance (measurable amount of light per given area)	115 cd/m^2	300 cd/m^2	*Not indicated*	*Not indicated*
Brightness (perceived amount of light dim, bright, very bright given certain conditions)	400 cd/m^2	300 cd/m^2	Variable (responds to sensors) and adjustable	Variable (responds to sensors) and adjustable
Power consumption: watts	150W max on	40W on, 2W standby	Up to 10 hrs when fully charged	Up to 6 hrs when fully charged
Viewing angle: max angle at which images on display can be viewed in degrees	178° (horizontal) 178° (vertical)	160° (horizontal) 160° (vertical)	178°	*Not indicated*
Weight: dependent upon technology used (CRT much heavier than LCD)	27.5 lbs	10.1 lbs	1.5 – 1.6 lbs	4.05 ounces
Viewing distance	24”	24”	16”	12”
Cost: dependent upon technology being used	$1799	$140	$499	$199

Factors: Resolution, Visual Acuity, and Field of View

It has long been thought that the most important factor in display quality was resolution. Early displays were riddled with communications issues like static, interference and signal interruption. The cathode ray tube is now known more for its sheer size and weight than for anything else. As technologies have changed, display profiles have trimmed and optimal resolution has become a moving target. LCD and plasma displays now tout incredible resolutions but each come with tradeoffs in achieving this. More and more we move into the realm of emulating reality (and beyond) through displays and the advent of three dimensional display technologies is upon us.

But how much does resolution really matter? What about luminance, aspect ratio, brightness, contrast ratio, viewing angle, refresh rate and response time? Are there minimum standards which should apply to the manufacture of all displays? Beyond readability and accurate representation are notions like immersion and presence. Some researchers suggest that displays have evolved beyond the limits of human visual acuity (Raghunath, et al., 2004). Others suggest that resolution and visual acuity don’t matter as much as field of view for quality of experience and degree of immersion (Lin, Duh, Parker, Abi-Rached, and Furness, 2002). In the field of virtual reality, it could be argued that feelings of immersion are linked to interactivity (Hwang et al., 2006)

According to Raghunath, et al. (2004), “Given that even people with perfect vision cannot resolve details smaller than one minute of visual-arc angle, increasing display resolution beyond that point does not contribute significantly to improvements in the amount of information shown.” With the advent of iPhone 4 and the “Retina Display” (http://www.apple.com/iphone/features/retina-display.html), even mobile devices with their very small display sizes, are claiming to have maximized human viewing capacity (at least for a certain viewing distance) by packing more pixels per inch (ppi, 326 for 4G and 130 for 3G). An improvement in the viewing angle and increase the contrast ratio are also enhancements in iPhone 4 display capabilities. In the table below, the relative resolution of a variety of display devices is presented along with the maximum possible display resolution according to human limitations.

Table 2. From Raghunath, Narayanaswami, and Pinhanez, 2004.
For up to date display resolutions see http://en.wikipedia.org/wiki/Display_resolution.

Table 1. Current and maximum possible display resolution according to human limitations.

Display type	User distance in inches	Typical width in inches	Typical width in pixels	Typical resolution in DPI	Maximum resolution* in DPI	Maximum width in pixels
Cell phone panel	10	1	100	100	350	350
PDA display	12	2	300	150	291	582
Laptop display	16	10	1,200	120	218	2,180
Desktop monitor	20	15	2,000	133	175	2,625
Laser printer hardcopy	12	7	2,100	300	291	2,037
Television set	100	25 **	694 **	28	35	805
Meeting room screen	230	80	1,200	15	15	1,200
Movie screen	500	720	5,000 ***	7	7	5,040

* Maximum resolution is computed using 20/20 human visual acuity, which is one minute of arc.
** For television sets, resolution is computed in the vertical considering the NTSC limit of 520 lines.
*** Considering standard 35-mm film stock with 4,000-dpi grain.

Deering (1998) adds that a typical CRT is as immersive (in terms of resolution and FOV) as a head mounted display device. Note that the human eye's optimal FOV as noted in the table can be 'saturated' by some of the visual configurations currently being produced. As Deering states, due to advancements in visualization technologies, particularly as 3D graphics, frame rates and resolution improve "The ultimate limits of human visual perception must now be included in hardware trade-offs."

In his Information Visualization (2000) text, Ware asserts that a 4000 x 4000 display (16 million pixels of a standard display size and distance to surface) should be adequate for any visual task based on the "resolving power of the human retina in each direction." In a recent study by Yost, Haciahmetoglu, and North (2007), there is strong evidence that large displays with increased amounts of visual information do not reduce accuracy and potentially improve it for certain types of tasks, while in some cases causing increases (3x) in task completion times. Display size is closely related to the interaction techniques provided for spatial navigation. If the display size is too small, additional navigation (pan and zoom) is required by the user. Better techniques for panning (finger drag) and zooming (pinch/expand) allow smaller displays to function closer to the performance of big displays, while poorer interfaces cause more of a difference (Hemminger, 1992.).

Hemminger (1992) also demonstrates that interaction styles may vary in insufficient and sufficient display size situations. Mental model selection for these two conditions can be critical to reducing cognitive overhead. For example, the 'filmstrip' style of interaction, moving images across the screen horizontally, may be appropriate when there is sufficient display size for the task, no navigational overview is needed. When the display size is insufficient, being able to zoom in and out of the image for greater detail or overview is a more successful method of interaction requiring minimal cognitive load. In addition to this, certain tasks may require specialized settings to optimize viewing using a given display.

Despite interest in larger displays, more emphasis in recent years has been placed on small displays and their strengths and weaknesses. Despite heralding the convenience of mobile devices, the fact that they fit in the palm of the hand are easy to carry and weigh very little, a tension remains between their reliance on battery power and their single biggest consumer of power, the display (Capin, Pulli, and Akenine-Moeller, 2008). Because of their portability, they can be used in a variety of contexts with variable lighting conditions which only adds complexity to the challenges associated with using them.

Like desktop systems, early mobile phones were large pieces of equipment that required a constant connection to power. When they moved off of the desktop into the car, they were still very cumbersome. The confluence of analog (1G) cellular phone system, small display technology for consumer electronics and improvements in battery technologies resulted in a 1973 Motorola prototype mobile phone. Early mobile phones were modeled on the typical keypad of a telephone and incorporated a very small display with a huge integrated battery. These early devices though not very ergonomic, provided an opportunity to send and receive calls without wires.

Comparative Studies

Some of the most interesting work looking at comparisons between large and small displays is taking place in medicine. Beard et al. (1993) compared radiologists' review of images on a display versus that of a conventional horizontal film alternator. Findings suggested that the computer workstation with a 2048 x 2560 pixel high-brightness monitor provided a clinical equivalent to the film alternator for reading chest CTs. Further work established that two 1024x1024 displays could perform as well for specific radiological reading tasks (Beard, Brown, Hemminger, and Misra 1991). Though visualization of radiologic images using mobile devices is an interesting area of research, display size limitations have inhibited progress. In Toomey et al. (2010), it is apparent that for certain types of radiologic imaging tasks, a mobile device may perform adequately. Similarly, even data intensive applications like GIS mapping, mobile devices may perform as well as other devices for certain types of tasks, and may provide added mobility for others (Clegg et al., 2006).

Jones, Jones and Dey (2004) conducted a study investigating the use of keyphrases, particularly when metadata is not available, as search result surrogates for small screen devices. In testing the keyphrase surrogate against a title surrogate among users of a small screen device, the authors found that categorization was roughly equal for each type of surrogate. What is more important, perhaps, is that the keyphrase surrogate can be especially helpful in the absence of good metadata or in cases where a title is poorly constructed or highly domain-specific.

Early work done by Jones et al. (1999), identified that in order to achieve a similar experience web browsing on a device with a much smaller display, navigation elements would need to change as well as interaction modality. This need only increases with task complexity. Though Capin et al. (2008) point out that mobile device limitations and their resulting displays and graphics are still limited by power supply, computational power, physical display size and input modalities, only physical display size and input modalities are expected to remain challenging in the years ahead.

As the cost of very large, very high resolution displays drops, their use increases and research into productivity benefits from the increased display area is ongoing (Czerwinski et al., 2003). At the same time research into alternative display forms such as electronic paper are also being investigated (Rogers et al., 2001). Meanwhile, mobile device display technology innovates on the theme of multi-touch devices with a wide variety of offerings which use differing technologies to improve the sensitivity and tactile feedback of these displays (Moscovich, 2007, Elezovic, 2008).

Solutions: Images

In addition to these elements of display quality are also elements related to presentation on a given display. Many techniques are invariably employed to overcome the problem of lack of space when using devices with small displays. The following elements become central to usability concerns as display size gets smaller:

- **Presentation**: what is presented and what isn't (e.g. peephole displays, "halo" etc.) and how
- **Text/Reading**: line length, text splitting, guided scrolling, RSVP, etc.
- **Interaction**: methods users use to interact with the system

- **Navigation**: features added to improve movement within and among visual elements
- **Design**: use of color, shape and layering to improve intuitive interaction with perceptual layers

Early work focusing on display size looked at text presentation. Still there is a great deal of effort going into how people read text and how best to present it on a display. From testing text presentation on paper versus that on a display to line length (Bernard, Fernandez and Hull, 2002) and the use of Rapid Serial Visual Presentation (RSVP) (Bernard, Chaparro and Russell, 2000) and beyond, HCI and cognition researchers have made significant contributions to understanding what makes a readable display. The heterogeneity of display sizes alone has made ways in which to interface with them more challenging.

Perhaps the real challenge/opportunity comes in the area of image manipulation. A great deal of work has been done on digital displays and image quality in the medical arena and in viewing radiologic images in particular. In Hemminger, Bauers and Yang (2008), emphasis was placed on comparing navigation techniques for large digital images. Five interaction styles were investigated: scroll bar, mag lens, pointer, arrow key and sectional (see Table below for more details). They found that the pointer interaction was preferred over all others and was described by subjects as being most intuitive and "mimic-ed net searching." In studies of 3D models and medical images, and their rendering on mobile devices, a central question revolves around the detail required for the task at hand, and, presumably, for the task 'in the field.'

Lots of work is being done to investigate support of remote health care work and particularly the transmission of high quality radiologic information (Andrade, Wangenheim and Bortoluzzi, 2003). This focus on data transmission and/or manipulation (Tang, Law, Lee and Chan, 2004) of image information rather than of visualization has met with some success (Correa, Ishikani, Ziviani, and Faria, 2008; Toomey et al., 2010). Perhaps more interesting is work involving diagnostic algorithms to support the data collection process in the field and to enhance the medical process by reducing time required for transmission and interpretation (Correa et al., 2008).

There have been many different approaches to handling large image navigation issues in the desktop environment. This is even more of a factor among mobile devices. Early work done by Yee (2003) on Peephole displays demonstrated an approach that would allow the user to view the context of the information space in an offset superimposed image while at the same time taking action in the main display window. Study findings yielded significant improvements in one-handed tasks using the Peephole display.

Baudisch and Rosenholtz (2003) introduced their 'Halo' concept which helps the user infer the locations of off screen objects with portions of onscreen 'rings' (visual references to off screen objects), thereby increasing the visual spatial range of a small display. For users of mobile devices, one of the biggest issues with context is display size. Because large images, maps, web pages, etc. are viewable using small screens and restructuring the information space is not always an option, a great deal of work has gone into optimizing interaction for viewing and interacting with them.

Jones et al. (2005) investigated speed-dependent automatic zooming (SDAZ), which combines panning and zooming into a single operation, on small displays. Recommended for improved navigation of large images on desktop systems, SDAZ was presumed to also be effective with small display devices. Their results suggest that, despite requiring fewer actions, use of SDAZ was not faster than using the standard interaction for tasks and that target acquisition was not more accurate. Subjects performed better on map tasks using SDAZ than on document tasks.

Chittaro et al. (2007) conducted an experiment comparing three techniques for navigating large information spaces (maps, webpages). They compared the use of a DoubleScrollbar (standard combination of two scrollbars for separate horizontal and vertical scrolling with zoom buttons to change the scale of the information space), Grab&Drag which enables users to navigate directly dragging the currently displayed portion of the information space with zooming handled through a slider control and Zoom-Enhanced Navigator (ZEN) which is an extension and adaptation to mobile screens of Overview&Detail approaches, which are based on displaying an overview of the information space together with a detail view of a portion of that space. Their findings suggest that factors like interactive update, sequential versus non-sequential navigation, navigation parameters, orientation cues, and user workload all play an important role in selection and preference of navigation techniques.

In his work on mobile visualization design, Chittaro (2006) suggests the following six steps when creating visualization designs for the mobile environment: mapping, selection, presentation, interactivity, human factors and evaluation. Chittaro also suggests that the traditional desktop solutions to presentation problems, overview+detail and focus+context, do not work in mobile environments. Instead, references to off screen information and more intuitive navigation techniques are required. Sensors that provide context or physiological awareness integrated into devices, particularly mobile devices, can provide enhanced information access that may supplement what is not achievable in a small sized display (Chittaro, 2008).

Another important newer area of research is that of immersive or virtual reality environments. What is significant in these environments is a change on the level of interaction which enhances the interface between computing device and human being. Hwang, Joong and Kim (2006) conducted a study comparing perceived field of view among a variety of display sizes with sense of immersion and presence. Their findings suggest that given a level of interaction with the device that involves motion, the perceived field of view with a small (handheld) device was much greater (50%) than actual.

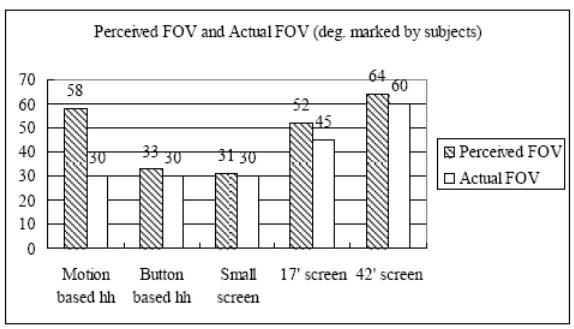

Figure 6. Perceived FOV (marked by subjects). Left is the perceived and the right is the actual.

Figure 3. From Hwang, Joong and Kim (2006)

Solutions: Search

Another area where mobile devices are increasingly being used is in searching tasks. Display size plays an important role in user experience when performing a search. A great deal of work is underway to find ways to improve the searching experience on mobile devices. Due to small size, there is an inextricable relationship between the interaction modality of a mobile device and its display. Many device manufacturers have tried to overcome this hurdle with specialized buttons, tailored browsers, and a variety of other tools including motion tracking.

Chae and Kim (2004) studied the important relationship between display size, task complexity and information structure. Their findings supported earlier work suggesting that the horizontal depth of information structures was a key element in the perception of task complexity. The specific challenges inherent with small displays for search are also being investigated. From visual snippets

(Teevan, Cutrell, Fisher, Drucker, Ramos, Andre and Hu, 2009) to keyphrases as surrogates (Jones, Jones, and Deo, 2004), the problem of reviewing and prioritizing search results efficiently is a critical element of search success.

More than this is the issue of marking up content, improving software applications and tailoring user experiences to a specific device. Layers are being created at every level of development. Low-level interaction devices vary from the use of a fingertip to a stylus to voice recognition. Operating systems vary; MacOS has hooks for multi-touch interaction that Windows has yet to employ. Software applications may exist in a variety of formats for specialized use on different devices from iPhone to Android to Symbian andPalmOS, etc. Web browsers specific for the device may interpret content with lists or navigation elements to improve user experience on very small displays. Content providers may 'sniff out' the device accessing a portal and serve up 'mobile' content designed for small screen interaction. Even content designers work to lay out and mark up their content in such a way as to provide the best experience possible, sometimes duplicating efforts to try and provide more comprehensive access.

To the user, sometimes these are real and vast improvements in user experience of a given website. It may take half the time to browse for a product given smaller load times for images, and improved navigation elements. In many instances, it can mean relearning the layout of an otherwise familiar 'full site or classic site'. For the mobile user, it can mean truncated menus, lack of facets, and awkward views of otherwise familiar content.

Site owners are drowning in an effort to give their customers options, trying to uncover ways to provide more and better support without increasing the cost to build and maintain the technology they need. In a study looking at J2ME cross platform deployment by Buranatrived and Vickers (2004) concluded that while it may be possible to deploy to multiple platforms, taking individual interaction differences into account would improve usability. For the developer, this may mean writing an application three times/different ways or more to improve access for the audience(s). To software engineers this may mean greater job security (more work) and a lot of reverse engineering. For the content specialist, this means considering every possible way a person might access a site today and in the future and building in a lot of redundant ways to access the same information or similar information.

Some publishers have already taken the leap in re-engineering their backend systems to help facilitate and manage content creation for multiple platforms. Thomson in particular spent millions of dollars recreating their authoring and editing system in order to dynamically re-construct or construct documents on the fly. To do this, they had to distill discrete components of information, like drug pricing information which changes rapidly, and manage them independent of any given publication. This allows them to update the price of a drug in one place and populate that change throughout their system, including any relevant publications which contain that information.

4. INTERACTION

In their 2009 article on Google search users of different devices, Kamvar et al. suggest that a close integration of mobile devices with the computer based interface would be beneficial to the end user because they treat these devices as an extension of their computer. But bringing this in line with the constraints imposed by the small form factor of these devices has been challenging. In their First Quarter 2010 report, Canalys noted that for the first time more touch screen smartphones were shipped than non-touch screen smartphones. It is no coincidence that touch screen technology has really hit the mass market in a mobile device, where its value in ease of operation may be felt the most.

User interface designers, computer scientists and programmers are struggling to develop application solutions which cater to the plethora of technology devices in use today. Though many of the so-called 'converged devices' or 'smartphones' are evaluated according to their feature set, few are ever used at the level of their real capability. To many in the information science arena, burgeoning mobile device use has created a perfect storm of sorts: where the need for very natural, easy to use interfaces has finally won out over the traditional keyboard, mouse and display paradigm, and where the promise of access is actually realizable for the bottom billion, in places without running water and often without paved roads. While it seems remarkable that these two things could coalesce, it seems very elemental that the three factors at work to produce this environment include a need for mobile solutions (ones that travel to or with the user), the ability to provide solutions anywhere, anytime through a wireless infrastructure, an interface that requires little to no training and/or prior exposure and voice communication as a central component (many in the bottom billion still haven't benefited from literacy efforts). Add to this an infrastructure that quite literally touches the ground lightly and you have a recipe for real and global change through technology.

Table 3. From Tarasewich, 2002

Table 1 – Wireless Device Categories
Laptop Computer
Handheld (e.g., Palm, Pocket PC, Blackberry)
Telephone
Hybrid (e.g. "smartphone" PDA/telephone combination)
Wearable (e.g., jewelry, watches, clothing)
Vehicle Mounted (in automobiles, boats, and airplanes)
Specialty (e.g., the now defunct Modo)

Still, end users are grappling with learning to use these technologies and their sometimes unique interfaces. Inherent in the challenge of multiplatform user interface design is the dilemma of designing for dynamic characteristics like task expansion, experienced users, and specialized or tailored feature sets. In assessing impact and spearheading change, information scientists struggle to focus research efforts on a set of discretely analyzable consequential tasks.

For decades now, the interaction paradigm has been focused on displays for output and the keyboard and mouse for input. Development of the Apple NewtonOS began in 1987 and included handwriting recognition which was implemented poorly in the initial phases. When the PDA products running NewtonOS hit the market, the problems had been improved. Some herald the development of the Apple NewtonOS and the corresponding line of PDA products the Message Pad and eMate in the late 1980s and early 1990s as the first big step toward mobile device computing with a new interaction paradigm. However, it is now generally understood that the Apple Newton was way ahead of its time.

When the uptake of mobile devices really took off in the early 2000s, (with several starts and stops in the preceding years) the stylus was re-introduced with Graffiti (unistroke gestures) for the PalmPilot and met with some success. Around that same time, the Apple iPod was released and the thumb wheel interaction became (and remains) popular. After beginning with two way pagers, the RIM BlackBerry PDA was introduced and the "thumbing" interaction took off. Though not the first to integrate a small keyboard into a PDA device, RIM's 'always on' or 'push' technology made these devices hugely popular in the business enterprise, enabling employees to "stay in touch". Today there are quite a number of keyboard variants from tilt keyboards to virtual keyboards with an almost equal assortment of sizes and layouts, most of which are variants of the QWERTY design based on early typewriter mechanical constraints.

Traditional input devices have included the keyboard, mouse, joystick or game controller, scanner, and a camera (still and video). Output devices have largely been limited to a display and a printer. All input devices include a sensor of some type to record movement, some method of providing feedback to the user about their movement, design features for ergonomic appeal, and interaction techniques that support the completion of a task (Hinckley, 2002). Most mobile devices of today may include a keyboard, stylus and/or touchpad, a multi-touch screen/display, thumbwheel or scrollwheel, softkeys and buttons, microphone for voice recording, camera, and internal sensors for orientation (gyroscope), acceleration (accelerometer), proximity and ambient light. Location based services (LBS) are also now integral to most mobile devices and can include GPS or assisted GPS, a digital compass, Wi-Fi, and cellular network assisted LBS. Output devices for mobile include sound, display, and sometimes other integrated devices.

In their 1993 article, Robert Jacob, John Leggett, John Myers and Randy Pausch, describe an *interaction device* as any device used to interact with a computing system, an *interaction technique* as the ways in which a particular input/output device is used to perform a task (Foley, 1990), and *interaction style* as any of the numerous ways the user can interact with the system. In an effort to further refine interaction style, Sharp, Rogers and Preece and (2007) and Shniederman, Plaisant, Cohen and Jacobs (2009) and Jacob et al. (1993) suggest that most styles fit into the following categories: command language, natural language, form fill in, menu selection, direct manipulation, and virtual reality.

Direct manipulation, popularized by the Apple Macintosh windows, icons, menus and pointers or WIMP interface, represented an important departure from the command line interfaces preceding this. Ben Shneiderman (1983) expressed the important attributes of direct manipulation as follows:

- "An object that is of interest to the user should be continuously **visible** in the form of a graphical representation on the screen
- Operations on objects should involve **physical actions** (using a pointing device to manipulate the graphical representation) instead of commands with complex syntax
- The actions that the user makes should be **rapid**, should offer **incremental** changes over the previous situation, and should be **reversible**
- The **effect** of actions should immediately be visible, so that the user knows what has happened
- There should be a modest **set of commands** doing everything that a novice might need, but it should be possible to expand these, gaining access to more functions as the user develops expertise."

Though seemingly obvious now that GUIs are the norm, it is not always possible to conform to this list of attributes, particularly as device form factor gets smaller.

Interaction devices can be direct (in sync with what is on the screen) or indirect (a representation) but occlusion of the field of view can be an issue with direct devices, an important concern with display-based multi-touch interactions. Interaction with the mouse as a pointing device has proven to be quite robust and in many types of fine grain tasks, outperforms direct manipulation (Barnert, 2005). In pointing tasks direct manipulation appears to consistently outperform (in both speed and accuracy) indirect manipulation (Kin, Agrwala and DeRose, 2009). Another factor in the indirect versus direct manipulation debate is the issue of hand and arm fatigue. These are particularly significant factors when multi-touch interactions are used with large displays (Wang and Ren, 2009). Both orientation of the display and types of interactions can play a role in this.

Despite the claim that "Electronic devices can become our eyes and ears and even our arms and legs" (Clausen, 2009), the vast majority of computer users still interact with the system using devices that have been around and improved upon over the last several decades. As Bill Buxton (2008) suggested, interaction device innovation has been moving at the speed of a glacier, until recently.

Design

Elements of good design tend to be simple. According to Sharp et al., (2007) three essential steps in the interaction design process include focus on the user, identifying specific usability criteria, and iteration of the design. Compared with designing for the traditional desktop system where interaction devices are somewhat constrained and well tested, designing for interaction with mobile devices is complex. There are currently on the order of 4,000 different mobile device models (not including the non-voice communication devices), nearly 200 manufacturers and half a dozen major operating systems (OS).

While issues of bandwidth, battery life, operating system and network availability aren't completely a 'thing of the past', they are largely addressable and can be expected to diminish as factors in designing for mobile devices. As the global development of wireless infrastructure expands so to do the possibilities of what can be accomplished with mobile devices. What will remain as major obstacles are display size and interaction style (Raghunath et al., 2003).

As A. R. Wilson (1998) so aptly noted in "The hand: How it's used to shape brain, language, and human culture", "touching, holding, and moving physical objects is the foundation of the long evolution of tool use in the human species." It is not hard to understand why the highest goal of interface designers is to model a device which is intuitive to users and ergonomically appealing. Yet given the constraints of mobile devices, this is often hard to achieve.

The morphological characteristics of mobile devices come from various origins. Characteristics from the Alexander Graham Bell telephone to the timepiece can be found in elements of modern mobile devices. Typical form factors of a modern mobile device include the flip/clamshell, dual hinge clamshell, candy bar, swivel, slider and slate. Whatever the form factor, the two biggest components of design are the display and the interaction device(s).

Display size for mobile devices can vary considerably. According to Tarasewicz, "Most mobile phones have small (1" to 2" square) screens that can display 4 to 8 lines of 10 to 20 alphanumeric characters each." The resolution range for today's mobile phone market is from 96 x 65 pixels (Nokia 7110) to 1024 x 480 pixels in a 3.8" display (Softbank 931SH). Apple's iPad includes a 1024 x 768 pixel display over 9.7". The first challenge for designers is to improve output for very small displays, the second is to enable human interaction with small devices and the third is to do this across the spectrum of devices available for a wide set of tasks.

According to Tarasewich, "many Web sites are trying to duplicate their wired Web architecture and design for the wireless Web, resulting in poor navigation and information overload." To help address this, Chae and Kim (2004) propose information structures with efficient depth and breadth in design. Depth is defined as the number of levels in the hierarchy and breadth is the number of options per menu panel . In their article "Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones," they report on a study investigating the relationship between screen size and task complexity using mobile devices with very small screens. Their results suggest that with simple tasks, the effect of screen size and horizontal depth are less significant than with more sophisticated tasks. For designers, the implications are that limiting horizontal depth for users of small screens may result in better navigation and an improved perception of usability.

In their article "Simplicity in Interaction Design," Chang, Gouldstone, Zigelbaum and Ishii (2007) define featuritis as "the tendency for designers to emphasize the number or novelty of features over core usability". This tendency, they note, is directly at odds with the users need for more explicit feedback given the increased complexity of devices. When their students were asked to impose design constraints that focused on the most simple, straightforward approach to design, they arrived at more designs which involved usage metaphors. Their conclusion: that simplicity could foster novel innovation in interface design.

Jones et al. (1999) provided an important design contribution in their study of task complexity and screen size. While the authors suspected that orientation of the small screen user on the content page would require a lot of back and forth scrolling, they observed only a lot of scrolling down and to the right to navigate content. For designers, their recommendations included providing direct access to content, providing direct search features, provide focused navigation by structuring information and make efforts to reduce the scrolling required by the user. Pettinati (2007) also recommends streamlining common use cases, exposing hierarchy and importance, display features progressively, highlight (enlarge) interactions that are actionable, make certain types of content actionable (phone numbers), design for the display (device specific CSS), consider device-specific interaction devices and network latency possibilities carefully.

Context

Though not a significant factor in interaction design for desktop use, the advent of wireless infrastructure has created a new critical element in design considerations, context. Context applies to both the physical location of the user in any given environment and to the nature of the interaction the user has through the display of the device. Context for the user given the limitations of the display was covered in the Display section of this review.

As Wobbrock (2006) points out, current trends in society and technology require that "the future of mobile HCI research be one which considers context as much as capability." Citing an increasingly aging population, the amount of computing work now done away from the desktop, the increased functionality of mobile devices and a general trend toward convergence of computing capabilities in a single device like the mobile phone, Wobbrock suggests that HCI research on mobile device has been limited to the device itself, focusing on facets like interaction, display size, browsing, domain specific applications, and so on, yet there is much more to be learned by going beyond the device.

Perhaps one of the greatest challenges in mobile device emergence, the inclusion of context, is also a great opportunity. In practical implementation, context may mean performing a Google search for local restaurants without having to include your present location. There is perhaps no other more ubiquitous piece of technology than the mobile phone. Thomas and Mahler paraphrasing Mark Weiser's 1991 article ""The computer in the 21st century", describe the future computing device as one in which "the computer should be integrated seamlessly, the user not being aware of its presence."

According to a review article by Dey and Abowd (1999), context may be described as "implicit situ- ational information." More specifically, it is "any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." Context includes information about the computing environment, the user environment and the physical environment. Certain elements of context are more important than others. Location, identity, activity and time are considered primary elements of context and are very practically important in context-awareness application development.

Context-aware computing is well situated to change the face of computing as we know it customizing applications to a user's current situation. Moreover, context enables providing task relevant computing services and information to a user (what they need when they need it), critical elements in the vision set forth by Weiser (1991).

Keyboard

Modeled after the typewriter, modern keyboard designs have retained an artifact of their early mechanistic challenge: the QWERTY layout. Resulting from a need to arrange the keys without overlap, the QWERTY layout also has the inherent advantage that it is more or less optimized for bimanual input. Leveraging procedural memory, QWERTY keyboards have been in use well over 100 years (Hinckley, 2002). Despite some potential efficiency and safety (less work-related strain) gains, the Dvorak layout has not garnered the same attention probably because of the overhead associated with retraining to use it (Hinckley). The keyboard has become the gold standard input device for text entry. "Although the ubiquitous QWERTY keyboard reigns supreme as the primary text entry device on desktop systems, mobile and handheld systems lack an equivalent dominant technology or technique for the same task." (MacKenzie and Soukoreff, 2002).

Despite the fact that usability testing has revealed the QWERTY layout to be suboptimal, it has remained the most commonly used basic keyboard for many years. This is due in part to its widespread adoption and the 'time' cost to change. There have been variations of the QWERTY keyboard, more so in recent years with the advent of mobile devices. In an effort to overcome the limitations of a small form factor, many different styles of keyboard and/or alternative methods of text entry have been employed. Some of the variants include full size QWERTY keyboard, mini QWERTY keyboard, multi-tap keypad (non-predictive), multi-tap keypad (predictive), soft keyboard (QWERTY, miniQWERTY, stylus based and multi-touch based) (From Curran, Woods and Riordan, 2004). Although nearly every variety of keyboard has been empirically evaluated in one way or another, the focus here will be on soft keyboards typically used with a stylus as they are well studied and likely to be used (in some variant) with mobile devices.

There has been a great deal of research on keyboard layout, keyboard design and keyboard use for certain types of tasks. There is a good deal of well substantiated theoretical groundwork on typing speed as it relates to task execution. In order to increase the display size of mobile devices while at the same time increasing functionality, a soft or virtual keyboard has been implemented. Research on these keyboards suggests that there are some important considerations to make in their design and implementation.

In their 1999 work on soft keyboard layout, Mackenzie and Zhang used a predictive model to evaluate low fidelity paper-based keyboard layouts to try and optimize one for longitudinal assessment. The study then compared the predicted model OPTI to the industry standard QWERTY. As with any experience computer user, testing OPTI required re-learning a keyboard layout. The model predicted that the OPTI layout would be about 35% faster than the QWERTY layout. After the initial learning curve, the OPTI layout did indeed outperform the QWERTY layout. By the 20th session, the WPM rate for the OPTI was 45 and for the QWERTY, 40. Not only was the typing speed faster with the OPTI layout but the error rate was lower.

In 2001, Mackenzie and Zhang published work on an empirical evaluation of the novice experience with soft keyboards. Their aim was to investigate size effects as well as random layout effects. A stylus-based soft keyboard setup was simulated using two different sizes of QWERTY layout and two different sizes of random key assignment layout. The objective was to understand the effect of size of keyboard (small, large) as well as the effect of keyboard layout (fixed, random) also considered to be novice user—in this case the novice user was simulated by random key assignment after every keypress. Consistent with what Fitts' Law would predict (see Methods section), there was no significant effect on size of keyboard. Keyboard layout effects, however, were significant. In addition, error rates were lower for the random layouts than for the QWERTY layouts—probably a result of the participant having to locate the target prior to selection each time. The behavior of the participants (hovering above the keyboard to identify the next key) supported rejecting the hypothesis that Fitts' Law could be used to predict novice user behavior with the random layout task.

Sears and Zha conducted a similar study in 2003 evaluating soft keyboards in three sizes: small, medium, and large. The evaluation included a two screen (abc, 123) QWERTY soft keyboard layout and stylus to perform six tasks of differing complexity. While there was no significant effect for keyboard size, there was a significant effect for task type. Data entry rates, error rates and user preferences were not affected by keyboard size. While there were no effects based on keyboard size, there were significant effects (reduction in data entry rates) related to having to switch between keyboard layouts to complete a task.

In their 2007 study evaluating the use of paper mockups for text entry using soft keyboards, Mackenzie and Read has at its focus determining whether a paper mockup can serve well for empirical investigation of soft keyboard layout. Incorporated into the study are some interesting design components including use of research subjects for data capture. Results of the study were compared with prior work and typing speeds were found to be along the lines of those measured in more rugged empirical analyses. The authors conclude that this inexpensive and low fidelity approach to data collection and soft keyboard testing proved to be "a quick and efficient means to empirically test soft keyboard layouts."

Text Entry

An easy way to determine a core aspect of usability of a system was to investigate how well certain types of tasks could be performed using the system. This meant that text entry, as measured by something like typing speed, was an early and robust corollary to usability. Many, many studies investigate the effects of text entry on computing devices and most of these are beyond the purview of this review. Since our focus is on mobile systems, our interest is in how comparable text entry speeds on mobile devices are to traditional computing environments.

In his Chapter on Evaluation of Text Entry Techniques from *Text entry systems: Mobility, accessibility, universality*, Mackenzie details some of the critical reasons for evaluation and testing of text entry techniques. According to him, too often, great ideas remain inadequately tested or untested, due in part to an unfortunate reluctance of researchers to engage the user community. The point driven home in this chapter is the need for comparative analysis of text entry systems and that in order to accomplish this, standards and methods must be adhered to. Following the mores of experimental psychology, Mackenzie argues, questions should be "repeatable", "observable" and "testable" (ROT).

Curran, Woods and Riordan (2004) conducted a study of novice, intermediate and expert users of mobile phones and asked the groups to use a keypad based phone and a non-standard keypad based phone as well as a stylus based PDA with both a mini soft keyboard and handwriting recognition being tested. The predictive text (T9) function of the keypad phones was used both turned on and off. In addition to these devices, a full size QWERTY keyboard and a mini-QWERTY keyboard based device were included in the testing. Their results showed that in both preference and performance, the full size QWERTY computer keyboard was the fastest means of text input. It was followed by the mini QWERTY keyboard then by the soft QWERTY keyboard. The predictive text entry method was generally quicker than non-predictive though prior experience predictive text entry might have been important. Their results provide some information stratified by gender and age and they include a detailed treatment of error. This study is particularly nice because it included a wide variety of devices as well as the baseline or gold standard device a full size QWERTY keyboard. Despite this, there were some limitations of the study in terms of generalizability due to small sample size, especially with stratification.

Myung (2004) looked at mobile phone text entry among Koreans. Pointing out that the keyboard layout for the Korean alphabet had not yet been adopted (culturally and/or nationally), part of the study was aimed at determining whether a predictive model could be used as an alternative to empirical analysis to determine best layout options. KLM-GOMS was used to predict usability of new keyboard/keypad layouts of the Korean alphabet and this was determined to be as effective as empirical validation of the new layout.

In 2001, Isokoski and Raisamo introduced their Minimal Device Independent Text Input Method (MDITIM). Intended to model device independent text input, this proof of concept was modeled on simplicity. To validate MDITIM, a study was conducted and text entry was compared using a variety of devices including stylus on touchpad, mouse, trackball, joystick and keyboard. Though this approach was (and still is) somewhat contrary to the trend toward task specific interaction devices and/or techniques, it was a new approach to measure the same technique across different devices. This served to highlight the fact that operationally, though the stroke might be the same for MDITIM in theory, it was executed differently on each input device.

Zhang, Li and Blumberg (2008) highlight some key design considerations related to reading text on small screen devices. One important note is that their work was focused on Chinese characters. Their results provide insight into ways in which designers for small display devices might optimize font style, size and color to improve readability and reduce fatigue.

Kamvar (2008) and Cox, Cairns, Walton and Lee (2008) both investigate instances where voice recognition is being used to provide an alternative to keyboard based text entry. Kamvar investigated the use profiles of users of the Google Mobile Application when the voice search function was invoked. Their aim was to understand when and why users chose to speak their queries. Results suggested, contrary to the researcher's initial thinking, that longer queries were not the focus of voice searching.

Cox et al. (2008) compared voice based text entry to multi-tap and predictive text entry to validate KLM predictions. They then investigated these text entry methods in limited visual feedback conditions to determine the value of voice based text entry under conditions like walking, driving, etc. Based on their predicted results, a combination of keypress and voice recognition would yield the best task completion time which was in fact the case. For more on this modality, see the voice section that follows.

Das and Stuerzlinger (2008) investigated an important area of text entry, learning effects. Their work resulted in a predictive model that could be tailored to user experience level, helping to elucidate the quantitative measures of learning effects (between novice and expert). This predictive model was tested against simulated users and was found to be highly accurate. Though empirical testing should be used to validate these results, the adjustments made to the model are informative for testing text entry among mobile phone users.

Pointing/Mousing

The advent of the mouse signaled a significant shift in human computer interaction. Made popular with the release of the Apple Macintosh, the mouse has undergone several transitions from a mechanical ball design to an optical mouse with fewer moving parts. Communication routes for the mouse have also shifted over the years from PS2 to USB and so on. More recently, the mouse has become untethered using various wireless protocols like Bluetooth to communicate with the computer system. The integration of buttons and of additional functionality like that of multi-touch capability (see Apple's Magic Mouse) has improved the functionality and usability of the mouse in recent years.

Use of the mouse as a pointing device has been well studied (Card, English, and Burr, 1978). The primary focus for quantitative evaluation of usability of the mouse has centered on the use of Fitts's Law (Fitts, 1954). An early comparative analysis conducted by Mackenzie, Sellen and Buxton (1991) investigated the performance of a mouse, a trackball and a stylus with a tablet in pointing and dragging tasks. Their results confirmed the work of Card et al. (1978) suggesting that the mouse performs well for pointing tasks and extended this to include the stylus and tablet which performed nearly as well. There were clear differences in performing pointing tasks and dragging tasks, the trackball performed poorly in both types of tasks. Their results also confirmed that Fitts' Law could be used to model both pointing and dragging tasks. They also suggest that the stylus tablet combination may be more suitable for finer pointing tasks such as drawing or gestures where the mouse performed best overall for dragging tasks.

Mackenzie and Isokoski (2007) evaluated throughput when performing a pointing task. Using Fitts' reciprocal tapping task, subjects were asked to complete a block of tapping tasks under three different conditions: normal, speed as a priority and accuracy as a priority. The goal of the study was to determine if throughput was affected by changes in cognitive focus resulting in different movement times and/or error rates. The results, helping to support Fitts' original premise that throughput would be constant, suggested that indeed regardless of cognitive focus, throughput remains the same.

In a 2009 article, Sasangohar, Mackenzie and Scott investigated differences between mouse and touch input for a tabletop display. Again, using Fitts' reciprocal tapping task, throughput, movement time and error rates were measured and compared. Touch interaction yielded a higher throughput than mouse interaction though with more errors for small targets. While survey data suggested that touch interaction was also preferred, small target selection is expected to remain a problem with touch based interaction.

Gestures and Multi-Touch

As Moscovich (2007) and Buxton (2008) point out, despite capabilities otherwise, much of our interaction with computing systems has been constrained to a trickle through a single-point input device. [The] "Multi Touch User Interface is a multifunctional gestural interface using hardware and software to recognize, track and interpret multiple simultaneous touches on a touch screen." (Elezovic, 2008). Saffer (2009) refines this further and describes actions performed with touchscreens and interactive surfaces as including: Tap to Open/Activate, Tap to Select, Drag to Move Object, Slide to Scroll, Spin to Scroll, Slide and Hold for Continuous Scroll, Flick to Nudge, Fling to Scroll, Tap to Stop, Pinch to Shrink and Spread to Enlarge, Two Fingers to Scroll, and Ghost Fingers. For free form interactive gestures, he includes: Proximity Activates/Deactivates, Move Body to Activate, Point to Select/Activate, Wave to Activate, Place Hands Inside to Activate, Rotate to Change State, Step to Activate, Shake to Change, and Tilt to Move. Citing Japanese product designer Naoto Fukasawa, Saffer suggests that developers follow the "dissolve in behavior" rule that allows the product to dissolve into the behavior of the user.

Karam and schraefel (2005) made an important contribution to the study of gestures in HCI by creating a classification system that broadly describes application domains, enabling technologies (both perceptual and non-perceptual), system response and gesture styles. Drawn from the literature, they describe gestures as falling into one of five categories: deictic, gesticulation, manipulation, semaphores and sign language. They distinguish between deictic, manipulation, semaphores and gesticulation as follows:

> **Deictic:** "pointing to establish the identity or spatial location of an object within the context of the application domain"
> **Manipulation:** "a manipulative gesture is one whose intended purpose is to control some entity by applying a tight relationship between the actual movements of the gesturing hand/arm with the entity being manipulated."
> **Semaphores:** "we define semaphoric gestures to be any gesturing system that employs a stylized dictionary of static or dynamic hand or arm gestures..."
> **Gesticulation:** "one of the most natural forms of gesturing and is commonly used in combination with conversational speech interfaces"
> **Language gestures**: "Gestures used for sign languages are often considered independent of other gesture styles since they are linguistically based and are performed using a series of individual signs or gestures that combine to form grammatical structures for conversational style interfaces."

Until recently, most touch screen implementations included a stylus as the device of interaction. While the stylus affords a great deal of precision, it is still an indirect instrument and less intuitive than gestural interaction involving the hand(s), for example. Furthermore, handwriting recognition is still significantly slower than other forms of gestural interaction, error prone and slower than traditional keyboarding. Optical Character Recognition (OCR) paired with a camera was once thought to be a great way to improve the desktop working environment though performance of the OCR systems and processing time have limited this option.

Myron Krueger, a pioneer in virtual reality, is often considered the father of modern multi-touch having created an artificial reality type interface in the 1970s which remains more sophisticated than most HCI interfaces today. Much of his work was used in military applications and was originally oriented toward interactive art. He is credited with originating the pinch grasp movement typically employed in map applications in the multi-touch environment of today.

Historical accounts credit Nimish Mehta for creating the first touch screen prototype while a student at the University of Toronto in 1982. The following year, researchers at Bell Labs published a document on multi-touch though a product never followed on to this. According to Bill Buxton (Surface and Tangible Computing, and the "Small" Matter of People and Design, 2008), after the Mehta prototype was completed, he (Buxton) saw a much better version at Bell Labs. "The problem was that they [Bell Labs] never released the technology, so, the whole multi-touch venture went dormant for 20 years (Buxton, 2008)."

Pierre Wellner introduced his DigitalDesk calculator in 1991 and a more comprehensive electronic office working environment in "Working with paper on the Digital Desk" in 1993. Different from prior work, Wellner attempted to bring electronic capabilities to traditional working environments. This was considered to be the opposite of simulated worlds and virtual reality and Wellner called it augmented reality (AR). One of the biggest strengths of early AR development was its human centered design approach, as much as possible, the simulated environment was created in synthesis with human movements.

Buxton and Myers (1986) completed some early work on bimanual input for continuous (such as pointing and dragging with a mouse) tasks. Their results suggested that users could engage in the completion of subtasks simultaneously (with different hands) and that this 'parallelism' suggested the cognitive overhead to complete the tasks was minimal. In addition, subjects who engaged in this parallel behavior were more likely to complete the tasks more quickly and outperformed the single handed task on several different measures. Despite this, not all tasks are equally well suited to bimanual input. In his 2008 chapter entitled "Two-Handed Input in Interaction", Buxton illustrates the artificiality of single handed input as a constraint of the current computing environment. Still, he argues, there are many basic tasks for which single handed input is still optimal. Moreover, most bimanual tasks are asymmetric, that is they require primary focus from one hand and secondary support from another. Though an important area of research, until recently, bimanual interaction has been very limited.

Lee, Buxton and Smith (1985) introduced one of the first multi-touch tablets. They described their work as innovative in two particular ways: "First, It [the tablet] can sense the degree of contact in a continuous manner. Second, it can sense the amount and location of a number of simultaneous points of contact." Though not the first touch sensitive tablet of its kind, this was the first prototype with these important characteristics of multi-touch interaction.

In important early work comparing architectural tasks (sketching and sorting) with different display sizes (tablet, typical monitor and digital desk) and interaction styles (stylus with touch screen, mouse), Elliott and Hearst (2000, 2002), found that interaction style, display size and task type were dependent upon each other. For sorting tasks involving a significant portion of the workstation, intermediate sized displays were preferred (errors tended to occur with items in the periphery). Both qualitative and quantitative measures were analyzed. For sketching tasks, stylus based input was preferred (over mouse-based) and tablet sized displays were suboptimal. Low resolution of the large display was not a significant factor but readability on the tablet could be. Quantitative analysis did not support the initial hypothesis that "architects would prefer completing image design tasks on the Digital Desk" (Elliott and Hearst, 2002). In fact, the Digital Desk was not preferred for the sorting task and only partly preferred for the sketching task.

In 2005, Jeff Han introduced the use of frustrated total internal reflection (FTIR) to produce high-resolution multi-touch sensing displays. This technology provides "full imaging touch information without occlusion or ambiguity issues." Future work will include proximity information and a classification (e.g. which finger) for each point of contact.

Large scale multi-touch displays tend to be used for collaborative work. Elezovic (2008) put together a low cost proof of concept multi-touch interactive whiteboard system using wiimotes (as HID compliant devices with internal infrared cameras), infrared pens and GlovePie. Both multi-touch and multi-person, this concept is highly scalable and cost effective.

Exploring the wide variability and "guessability" inherent in gestural interaction, Wobbrock, Morris and Wilson conducted a user-centered design experiment with a Microsoft Surface prototype. User defined gestures were compared with expert generated gesture sets and found to have only about 60.9% agreement. For the vast majority of referents (tasks), gestures involving only one hand were used and preferred. Gestures which were deemed to be more complex also rated more poorly in terms of goodness and ease. Cognitively complex referents were not necessarily associated with poorer ratings in terms of goodness and ease though planning time had an impact on the perception of goodness and ease.

In a 2008 experiment studying the use of physical edges to improve target acquisition on mobile touchscreens, Froelich, Wobbrock and Kane investigated the effectiveness of this approach among typical users as well as users with motor impairments. The motor impaired user has difficulty interacting with the latest generation of smartphones which utilize multi-touch based touchscreens and have few physical buttons. This study investigates the use of barrier pointing to overcome these limitations. Results suggested that for certain motor movement impairments, particularly those which extremely limit fine motor control, barrier pointing can be useful.

Sun and Hürst (2008) present video browsing techniques like the mobilezoomslider, scrollwheel and elasticslider. While there were no significant effects in performance when comparing the elasticslider with the traditional iPhone interaction, individual preferences for interaction type were polarized. Further evaluation of these techniques may yield helpful information for improving video navigation on small screen devices.

Hoggan, Brewster and Johnston (2008) investigated the importance of tactile feedback during touchscreen use. Comparing a physical keyboard, a touchscreen keyboard and a touchscreen keyboard with tactile feeback incorporated, they found that the addition of tactile feedback brought touchscreen text entry to performance levels near that of the physical keyboard. A second portion of this analysis determined that tactile feedback enhanced with actuators that could provide specific feedback (location where button press was activated) could improve performance even further.

5. SEARCH

A great deal of foundational work has been done in the area of information seeking and retrieval (Bates (1979), Wilson (1981), Belkin (1988), Dervin (1992), Marchionini, (1991), Kulthau (1993), Hsieh-Yee (1993), Wildemuth, (1995), Borlund and Ingwersen (1997), and many others). After laying some groundwork in what searching is, much early work in this area focused on information seeking in different contexts (e.g. professional) and domains (e.g. library and information science) centering on the concepts of task and goal as they relate to an information seeking activity. While much of this work focuses on the "who, what, when, where, how, and why" there is still a lot of interest in both the reason for the search and a relative measure of success when a search is undertaken. A recent shift in research methods to the use of transaction logs in examining web searching behavior (Rose and Levinson, 2004 and Jansen et al., 2009) has met with both success and criticism. While the logs may be exacting in what the user actually does in interacting with a system, researchers increasingly want to understand more about the cognitive mechanisms associated with search. Doing this requires an expansion of the notions in existing theory which center on quantitative approaches to cognition and new methods for capturing the details of a searching 'transaction.'

Information search is a central theme in information science and has been theorized about since its inception. The details of search are still elusive elements spurring further research in the field. How and when does an information need arise? What tactics does the user employ to conduct a search? What is the intent of the search? How successful is the search? Mobile devices have only added complexity to these questions by facilitating searching in any [mobile] context.

Bates is perhaps credited with initiating the discussion on search tactic. In an effort to understand and disseminate the skills of experienced information searchers, Bates articulated and named a series of models of search strategy, idealizing, representing, teaching and facilitating searching. She goes on to elucidate tactics employed as part of the overall search strategy: monitoring, file structure, search formulation and term tactics. To each tactic is then added a set of defining terms which should aid the user in the process of searching.

T.D. Wilson (1981) put forward a model of information behavior stressing three important components: exchange—that information seeking involves some type of reciprocity; failure—that the needs of the user may be met or not met; and use—that the information will be used regardless of whether or not the need was met. He also puts forward a model of the context of information seeking in a universe of information. The 'need' in information need Wilson suggests, implies a basic human need. He asks whether an information need is a physiological, cognitive or affective need and goes on to suggest a model for information needs and seeking. He concludes that perhaps it might be more appropriate to say that we are engaging in information seeking in order to satisfy needs and that the information need is secondary to a given primary need. Factual data, he suggests, might satisfy cognitive needs, the channel of communication might be guided by affective or cognitive needs and the physical document may satisfy and affective need or in rare cases, a physiological need. According to Wilson (1981), "the communication model proposed by Shannon, [22] with its elements: source, channel, message, coder, decoder, receiver and noise, was never intended as an information-science model nor as a behavioural science model, and, consequently, can tell us nothing about the information user and his needs." This ends in a strong call for interdisciplinary approaches to information science research focused on the user utilizing social research methods rather than the 'user studies' done to date.

Kuhlthau (1991) introduced The Information Search Process (ISP), a six-stage process that information seekers go through when seeking information. The six stages include Stage 1: Initiation, Stage 2: Selection, Stage 3: Exploration, Stage 4: Formulation, Stage 5: Collection, and Stage 6:

Presentation." Her work underscored the affective component of information seeking, an element that had not been fully recognized prior to the introduction of the ISP. Typically, early stages of the ISP involve some degree of uncertainty. The act of information seeking results in an effort to reduce uncertainty and can provide the catalyst which begins the information search process. Kuhlthau introduces the six corollaries of the principle of uncertainty: process, formulation, mood, redundancy, prediction and interest. Kuhlthau draws our attention to the 'zone of intervention' created by the increased uncertainty that spurns the information search process.

Introducing the concept of sense-making, Dervin (1992) suggests that information retrieval, for example, can be informed by predictions provided through sense-making theory. Sense making assumes that the way people see their gaps informs the way they try to bridge them and that the essential aspects of information use can be captured by looking at these gaps and bridges. Universal gap-definings, Dervin asserts, can be put into categories which include creating ideas, finding directions or ways to move, acquiring skills, getting support, getting motivated, getting connected to others, calming down or relaxing, getting pleasure or happiness, and reaching goals. Sense-making, by definition extremely individualistic, is also both quantitative and qualitative and thus, requires both theoretical approaches for research.

In a recent study of military family internet use, Freedman and Henderson (2008) sought to better understand the impact information access through technology has on an individual's ability to cope. They propose a model of sense-making that includes four steps: scanning, interpreting, taking action and ability to cope. Greater access to information does not necessarily mean improved ability to cope, but the need to cope often results in actions which leverage technology and information. Though access to both technology and information are socioeconomically distributed, or not equally available to all, efforts to bridge the digital divide by increasing digital access could be confounded by variability in ability to cope and resulting digital behavior. The results support the notion that increased access supports ability to cope though demographic characteristics were not responsible for digital behavior, the tendency to scan, interpret and act.

Marchionini and Komlodi (1991) disambiguate information search and information retrieval in an important way. While an information seeker may engage in information retrieval, "machines cannot engage in information seeking but machines can engage in information retrieval." They similarly provide clarification on browsing, information seeking and learning suggesting that each in turn require a greater and greater commitment on the part of the seeker in the search process.

In his Chapter Information and Information Seeking, Marchionini (1993) makes a critical point about information access in a digital world. "We are dependent on machines to express this information in forms that we can perceive." While it is true that technology and information are mutually reinforcing phenomena, it is also true that the linchpin of human computer interaction is the interface. (Marchionini and Komlodi, 1991).

In her article "Where should the person stop and the information search interface start?," Bates points out that a critical element in information search is the role of the user. Despite great advances in search system design and implementation, users perceive there is an implicit 'search system knows best' kind of approach to searching. Still, many users want control of the search including the ability to determine what does and does not get included in the search and why. Dervin suggests that two things be made explicit in the development of IR systems: "the degree of user vs. system involvement in the search, and (2) the size, or chunking, of activities; that is, how much and what type of activity the user should be able to direct the system to do at once."

As Xie and Cool point out in their 1998 article "Online Searching in Transition," as searching becomes increasingly sophisticated and web searching the dominant paradigm, learning about different ways to interact in the searching process becomes more important. Users want to retain control in the searching process yet fundamental interaction constraints may make this complicated and difficult. Most importantly, library and information science curricula must shift toward incorporating tools and techniques for interacting in this new environment.

A review article by Martzoukou (2005) helps put some challenges in web information seeking research into perspective. Research in this area must be approached from a holistic perspective considering cognitive, affective and physiological elements but has been methodologically inconsistent and often lacks quantitative validity and qualitative consistency. This limits the comparative strength and generalizability of results. Improvements in the 'realistic' quality of search tasks, appropriate sample size, direct observation of subjects and adherence to some methodological tenets would all be steps in the right direction.

Information need, search behavior, and intent

Wilson (2000) reiterated a well known issue in human computer interaction: that prior to the 1980s and still somewhat present today is an inherent question about how users interact with a system rather than a focus on the information need the system is intended to assist with. "The studies reported, as virtually all to this date, are concerned not so much with human aspects of information use, but with the use of information sources and systems," without taking the needs of the user into account. Wilson stresses the need to focus on the individual and his/her needs and to take an interdisciplinary approach toward research in this area.

In a study of third year medical students searching MEDLINE, Wildemuth and Moore (1995) compared the quality of searches as determined by the students themselves (self-evaluation), by librarians in four dimensions and librarians noted missed opportunities in the searches. A typical search involved fourteen statements, seven different terms and eleven results. Results indicated that student's searches were adequate according to the librarian's ratings (five point scale on all four dimensions). Self-evaluation also indicated that student's were satisfied with their searches. In terms of missed opportunities, 97% of the searches contained missed opportunities of some kind the most prominent of which was not using the controlled vocabulary (MeSH Subject Headings). One significant response to research like this has been that search systems cater to the incidental user, one who may have no knowledge of the syntax of the system.

Jansen, Booth and Spink (2008) used web search engine logs to derive a classification of user intent for web searching. Three classification areas, informational, navigational and transactional and their corresponding characteristics were then used to automatically classify web search log queries and measure the effectiveness of the classification. Applying the classification system to Dogpile search engine transaction log queries, an automated classification according to the system the authors developed was implemented. This was compared against prior literature and a group of manually classified queries. The automated system was found to be accurate for approximately 74% of queries. These data were based on a dataset of over a million and a half queries and though limited to a single web search engine log, the findings should be robust. Automated classification systems such as this could be used in real time to help developers analyze and improve upon their offerings by providing content directly suited to a user's query intent.

In his article "Exploratory Search: From Finding to Understanding," Marchionini (2006) points out that search is becoming increasingly sophisticated and that users who've grown up in a world where digital media is more or less native to them, will demand increasingly usable systems. A significant effect of large numbers of people engaging in exploratory search is the mining of data on search behavior to appeal to the user and/or engage in adversarial computing. As users move beyond finding to understanding in their searching behavior, Marchionini sees the advent of easy to apply searching tools to aid the user in this transition.

Search tactics, search task and search success

So, what impact does search experience and domain knowledge have on search tactics? Does prior search experience facilitate better search tactics? Does subject knowledge offer an advantage? Hsieh Yee investigated nine types of searching tactics divided into three categories: search term tactics, search monitoring tactics and search formulation and modification tactics. Term tactic variables included, the use of the searcher's own terms and the query language (OTAL), the searcher's reliance on the thesaurus structure for term suggestions (THAL), off-line efforts at term selection (PREP), online usage of search terms (ACT). The single search monitoring variable, CHECK, was the comparison of search question with a search in progress. The formulation and modification variables were inclusion of similar concepts or synonyms (PARALLEL), the tactic of finding similar items from a relevant item (TRACE), the searcher's combinations of search terms (MANIPUL) and the tactic of viewing records to find relevant items (BROWSE). Though the literature contains equivocal results on the effect of prior search experience (novice versus experienced), Hsieh-Yee found that "the two groups differed mainly in term selection, inclusion of synonyms, and manipulation of search terms." This became more evident when searching outside their subject area. So, while experienced searchers were able to compensate for lack of subject knowledge, "no matter which topic was searched, novice searchers displayed no difference in their use of search tactics selected for this study."

Wildemuth (2004) conducted a study of user search tactic formulation over time. Somewhat analogous to novice and experienced users, these subjects were medical students whose searches were recorded at three month intervals over a nine month period. All students were taking an introductory microbiology course and the searches recorded were directly relevant to the course. Findings suggested that most searchers engage in a gradual narrowing of the retrieved set in an effort to find the needed results. Another interesting finding in the study was that domain knowledge peaked during the course of the microbiology class and dropped off afterward. Assistance at each of these stages was integral to improving performance and persisted even after domain knowledge dropped off.

Byström and Järvelin (1994) undertook an empirical analysis of the relationship between task type and information needed for a task. Prior work had looked at the problem of task at the work or job task level and also had conducted the analysis after completion of the task. In this study, Byström and Järvelin broke down the tasks into discrete components and assigned complexity from the user's perspective, collecting data while the task was being performed. The following five task types were identified: automatic information processing tasks (a priori), normal information processing which require some case-based arbitration, normal decision where cased-based arbitration has a major role, known, genuine decision tasks where permanent procedures for performing the tasks have not yet emerged and genuine decision tasks which are unexpected, new, and unstructured. Information types needed in tasks included problem information, domain information and problem-solving information. In order to compute the task complexity level, an *Information Complexity Index* was devised from the information types used in the tasks. They found that "The contrast between simple vs. complex tasks underlines the importance and consequences of task complexity: In the latter, understanding, sense-making, and problem formulation are essential, and require different types and more complex types of information through somewhat different types of channels from different types of sources." This underscored the importance of both task complexity and information type in models of information seeking and use.

Jansen Booth and Smith (2009) employed Anderson and Krathwohl's taxonomy of the cognitive learning domain to classify searching tasks. The aim was to try and understand whether learning has important searching characteristics. Their findings suggested that searchers use searching primarily for fact checking and verification. For evaluating and creating information needs, searchers tend to rely on their own knowledge though different styles of learning can have a moderating effect on the searching process.

Xie (2009) explores the relationship between task type and the information search and retrieval process. Defining key dimensions of work tasks as nature, stages, and timeframe of the tasks, and key elements of searching tasks as origination, types, and flexibility, Xie analyzed information search and retrieval processes among workers in both a corporate and an academic setting. The results validated prior work suggesting that task drives the information retrieval process.

Query Formulation and Log Analysis

Web search logs have afforded an entirely new area of analysis of users and their search behaviors. Providing vast amounts of data on a huge number of users, these logs provide trace data that may help paint a picture about how people look for information on the web and whether they have success in finding it. Because the logs represent real data from real people (are naturalistic), they have even greater inherent value. However, these data are not always associated with demographic or other information about the user, and because they are trace data, only inferential conclusions about user behavior can be made. As Jansen and Spink (2005) point out, there is a "high degree of consistency at the session and query levels of analysis across multiple Web studies." In addition, these "similarities exist even with researchers studying various search engines and utilizing a variety of analytical methods, definitions, and metrics."

In reviewing a series (both related and unrelated to each other) of studies of web log analysis, (Jansen, Spink and Saracevic (2000), Spink, Wolfram, Jansen and Saracevic (2001), Spink, Wolfram, Jansen and Saracevic (2002), Jansen and Spink (2005)) important trends in research in this area are identified. Each study involved the use of web search transaction logs to identify trends in user query formulation behavior. The first study (Jansen, Spink and Saracevic (2000)), looked at a relatively small sample of users (at that time it was huge though) at a single point in time and recognized that few web search users were taking advantage of advanced search tools.

In 2001, Spink, Wolfram, Jansen and Saracevic analyzed over one million web queries posted by users of the Excite search engine and found that the "Number of queries posed on the Web is huge, but searching is a very low art." That is, users were using few search terms, viewing few web pages, didn't use advanced search features and make few modifications to their queries. Terms tended to focus on entertainment and recreation. The long tail of web queries was evident with a small number of terms being with high frequency and a large number of unique terms being used with low frequency.

Extending the previous study to include a longitudinal assessment of web queries, Spink, Wolfram, Jansen and Saracevic (2002) looked at data from 200,000 users of the Excite search engine in September of 1997, December of 1999 and May 2001. They saw a shift in term subject areas from entertainment and sex to commerce and people over the period despite query lengths and user frequency remaining roughly the same. Their findings suggested that either users needed to develop better searching skills or web search engines needed to improve the search interface, algorithms and relevant results. "An Excite results page contains 10 ranked Web sites, and the percent- age of Excite users who examined only one page of results per query increased from 28.6 percent in 1997 to 50.5 per- cent in 2001. By 2001, more than 70 percent of Excite users looked at two pages or fewer."

In a 2005 longitudinal web search transaction log study of European users of the popular AlltheWeb.com search engine, Jansen and Spink saw a decline in query length, and a decline in sexual and pornographic searching. Reviewing data on hundreds of thousands of users, they determined that only five or fewer documents were viewed by each user, spending only seconds per document. Nearly half of all documents were not topically relevant.

Jansen, Spink and Pedersen (2005) conducted a longitudinal analysis of AltaVista search engine queries to see how web searching behavior changes over time. This analysis is similar to others around the same period looking at web search logs to try and understand what user behavior patterns looked like at a single point and compared over time. Many important findings result: query and session length increased, term frequency decreased suggesting that queries become increasingly sophisticated as users gain familiarity and frequency of use increased. Because it would be hard to conduct such a study using other methods, one main contribution of studies of this type was to demonstrate that log analysis is a viable research approach. Studies of this type also add to a general body of research of web search log analysis with generalizable results (across web search sites).

In an effort to better understand what occurs in the query refinement process, Rieh and Xie (2006) collected web transaction log information on repeat users (6+ unique queries/session). Using a final set of 313 search sessions, they developed a classification for query reformulation with three top level facets: content, format and resource. Most query reformulations involve changes in content. Eight distinct modification sequence patterns were observed: specified, generalized, parallel, building-block, dynamic, multitasking, recurrent, and format reformulation. They conclude suggesting that "Multiple data collection methods (transaction logs, thinking aloud, interviews, etc.) can be employed to further explore the patterns of Web query reformulation."

Mobile Search

According to Church, Smyth, Cotter and Bradley in 2007 "It is likely that mobile phones will soon come to rival more traditional devices as the primary platform for information access." Though mobile searching is different from desktop searching and remains tied to task complexity, searching activities in the mobile environment increasingly represent and extend search in the desktop context. Because the user is mobile, contextual information can improve and enhance the search process, perhaps improving the overall searching experience. As mobile users become proficient, their internet browsing and search behavior expands. Limited only by display size and interaction device/style, users increasingly attempt to perform behaviors resembling those done in a stationary setting (Church & Smyth, 2009).

A 2006 study by Kamvar and Baluja analyzed web transaction log data from Google's mobile search sites. Over one million hits were included in the sample and examined for patterns of use. This represented the first large scale review of search data involving mobile phone access and the first ever from Google. A follow up study by the authors was conducted in 2007 and much of that work makes comparisons between the 2006 report and the 2007 data. The 2007 study also involved the analysis of over 1 million page view requests anonymized and randomly sampled from Google web transaction logs over a one month period in early 2007. Their findings included average mobile query length (2.56 words/terms) and an estimate of time to enter queries which was computed from the length of an entire transaction.

Table 4. From Kamvar and Baluja, 2006

Table 2. Summary of mobile search statistics in 2005 and 2007.

Mobile search statistics	2005	2007
Words per query	2.3	2.6
Characters per query	15.5	16.8
Percent of queries that had at least one click	<10.0	>50.0
Percent of queries that had at least one "more search results" request	8.5	10.4
Time to enter a query*	56.3	39.8
Time between receiving results and clicking on a spelling correction for a query*	15.6	15.1
Time between receiving results and clicking on a search result*	29.1	30.0

* Assuming 10-second network latency in 2005 and 5-second network latency in 2007

In fact the authors note, "Despite the drastically different input techniques used, the similarity in median and mean query terms across search mediums might suggest that the number of terms per query is currently a ground truth for today's Web search."

The time to complete a query was proportional to the length of a query. This is an interesting 'reverse computation' that should be verified in the field. Queries from PDA devices (typically equipped with QWERTY keyboards) were longer than queries from mobile phone however the time to enter a query on the PDA decreased by 30.1 seconds.

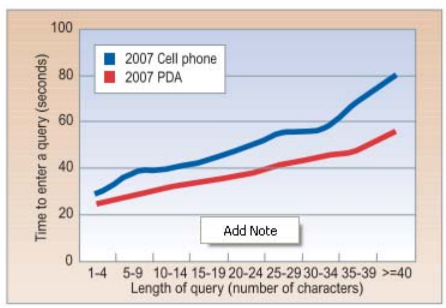

Figure 3. Time to query. Graph of the time it takes to enter a query versus the length of the query.

Figure 5. From Kamvar and Baluja, 2006

Top five categories or categories areas for the 2007 data are shown in the figure below:

Table 5. From Kamvar and Baluja, 2006

Table 1. The top five categories in mobile search.

Category	Percent of all queries
Adult	>25
Entertainment	>10
Internet/telecommunications	>4
Lifestyles/online communities	>4
Local	>4
Other	>45

The domination of the adult category is thought to be attributable to either the relative maturity of web search using mobile devices (a similar profile can be seen in desktop based web search) or to increased privacy on the mobile phone. Overall query diversity ranged from least diverse among cell phone users, next were PDA users and finally desktop users. Observing query pairs, the authors found that they tended to stay on topic and involve refinement. More than 50% of queries led to a click on a search result. The average number of queries per mobile session was found to be 2 and the time from Google front page to query submission decreased from 66.3 second in 2005 to 44.8 seconds in 2007, longer length queries saw a greater decline than ones of shorter length (see figure below). .

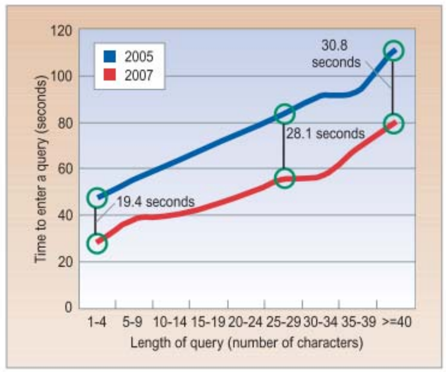

Figure 5. Reduction in query-entry time. Due to faster typing, it took less time in 2007 to enter a query than it did in 2005.

Figure 6. From Kamvar and Baluja, 2006

An increased interaction with search may be occurring in part because pages display better and also because interaction and network latency have improved. Overall, the number of queries and the query diversity per session increased though the query category appeared more stable. In general, queries are becoming less homogenous and the number of queries from cell phones and PDAs was evenly divided whereas PDAs took substantially less of a share than cell phones in the 2005 study.

Baeza-Yates, Dupret and Velasco (2007) make an important contribution in the area analyzing web transaction logs from Yahoo! Japan. One million mobile and one hundred thousand desktop **unique queries** created in 2006 were analyzed. Despite the fact that Japanese query terms are similar in length (2.3 terms on average), Japanese language differences (the use of characters) results in a substantial decline in average number of characters per query: 7.9 for mobile and 9.6 for desktop.

Comparisons with the Kamvar and Baluja (2006) study on query category are illustrated in the figure below:

Table 6. From Baeza-Yates, Dupret and Velasco, 2007

Category	Mobile	Desktop	Google	Category[5]
Business*	2.0	0.6	<2	Business
Business*	0.03	0.01	<2	Food & Drink
Business*	0.02	0.01	<2	Shopping & Consumer services
Games	4.6	8.0	>2	Games
Health	10.0	7.7	>2	Health & Beauty
Online shop	14.0	10.9	> 5	Internet & Telecom
Recreation*	5.6	3.6	>2	Travel & Recreation
Recreation*	0.3	0.1	<2	Automotive
Science	0.5	0.2	<2	Science
Sports	17.1	17.2	>2	Sports
Art	8.8	24.8	< 2	Arts & Literature
Computer	1.5	1.4	>2	Computers & Technology
Home	7.6	4.1	<2	Home & Garden
News	3.3	4.8	<2	News & Current Events
Recreation*	5.8	4.1	>10	Entertainment
Social	1.8	1.3	>2	Society

Table 6: Comparison with USA mobile search study (* = subcategories were used).

Because 'adult content' is not separated out in these categories, it is hard to know exactly how the datasets compare in that area. However, it does seem apparent that in Japan, a more mature region for mobile internet use, categories shift toward items more similar to that seen in desktop web search.

A study on European mobile users for both browsing and searching was conducted by Church, Smyth, Cotter and Bradley in 2007. Including more than 600,000 users, 400,000 query-based searches from more than 30 different mobile search engines, the data was collected over a 24 hour period in late 2005. This study particularly emphasized the difference between browsing and searching: 94% of all sessions were browsing sessions (following links) which left a fairly small subset (by comparison) for search analysis.

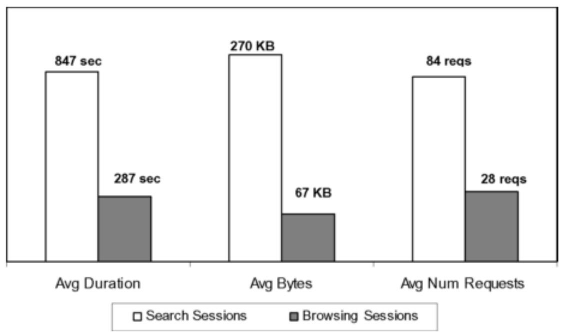

Fig. 2. A comparison between daily search and browsing sessions in terms of session duration, bytes downloaded, and numbers of requests.

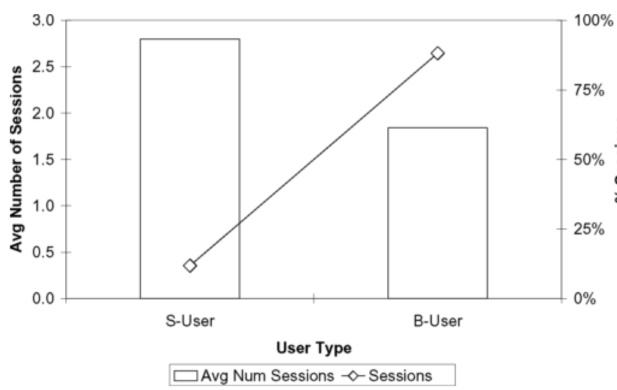

Fig. 4. Average number of sessions and percentage sessions per-S-User and B-User per-day.

Figures 7 and 8. From Church, Smyth, Cotter and Bradley, 2007

As the figures above indicate, the number of search sessions tended to be longer, more data intensive and include more interaction than the browsing sessions. In addition, the average number of sessions per search user was higher despite the total number of sessions for search users being much smaller. A critical element of this finding is that investigating search alone may tell an incomplete picture of how mobile users typically look for mobile information. This study is further differentiated by including multiple search engines.

Table 7. From Church, Smyth, Cotter and Bradley, 2007

Table I. Mobile Search Engine Usage in Order of Popularity

Rank	Search Engine	% of Searches
1	Google	76.2%
2	Operator-Specific Search	14.4%
3	Yahoo	5.7%
4	Independent Wap Search Engines/Directories *(Includes: TagTag, Click4Wap, Mooobl, Ithaki, Seek4Wap, Freowap, WapAll, WapMeter, etc)*	1.6%
5	eCommerce Search	1.6%
6	Other Major Search Engines *(Includes: AskJeeves, AllTheWeb, MSN, AOL, AltaVista, Excite, Nokia and HotBot)*	0.4%
7	Misc *(Includes: image-specific search, ringtone-specific search and other misc search services)*	0.1%

Important to note here is that mobile search terms were slightly less than that of web search (2.06 versus 2.3) and there was only minimal use of advance search features. This coupled with a lower number of unique queries, a higher incidence of repeat queries and more searches per session suggest that mobile search interfaces are insufficient for mobile searchers to locate information.

Despite the fact that the top 10 phones for browsing and searching were all standard mobile phones, important device characteristics for searchers included large screen size and higher resolution. Still most phones in the top 10 for each were quite modern with significant screen space, color interfaces, predictive input, and XHTML support. See the figures below for more details.

Tables 8 and 9. From Church, Smyth, Cotter and Bradley, 2007

Table II. Top-10 Mobile Devices Used in Browsing Sessions

Number	Mobile Device	% of Usage	Display Type	Display Size	Colors	WAP (Version)	XHTML	Email	Predictive Text	Navigation Control	Camera	Music Player	Radio	Video
1	A	9%	TFT	128 × 128	65K (16-bit)	2.0	x	x	x	x	x	x	x	x
2	B	7%	TFT	176 × 220	262K (18-bit)	2.0	x	x	x	x	x	x	-	x
3	C	6%	TFT	176 × 220	262K (18-bit)	2.0	x	x	x	x	x	x	x	x
4	D	4.1%	CSTN	128 × 128	4K (12-bit)	1.2.1	x	x	x	x	x	-	x	-
5	E	3.6%	CSTN	128 × 128	4K (12-bit)	1.2.1	x	x	x	x	-	-	-	-
6	F	3.1%	TFT	176 × 220	65K (16-bit)	2.0	x	x	x	x	x	x	x	x
7	G	2.9%	TFT	208 × 208	65K (16-bit)	2.0	x	x	x	x	x	x	x	x
8	H	2.7%	TFT	170 × 220	262K (18-bit)	2.0	x	x	x	x	x	-	-	x
9	I	2.3%	TFT	128 × 128	65K (16-bit)	2.0	x	x	x	x	x	-	-	x
10	J	2.2%	TFT	128 × 160	65K (16-bit)	2.0	x	x	x	x	x	-	-	-

Table III. Top-10 Mobile Devices Used in Search Sessions

Number	Mobile Device	% of Usage	Display Type	Display Size	Colors	WAP (Version)	XHTML	Email	Predictive Text	Navigation Control	Camera	Music Player	Radio	Video
1	C	17.6%	TFT	176 × 220	262K (18-bit)	2.0	x	x	x	x	x	x	x	x
2	F	7.4%	TFT	176 × 220	65K (16-bit)	2.0	x	x	x	x	x	x	x	x
3	B	6.5%	TFT	176 × 220	262K (18-bit)	2.0	x	x	x	x	x	x	-	x
4	A	5.9%	TFT	128 × 128	65K (16-bit)	2.0	x	x	x	x	x	x	x	x
5	K	4.8%	TFT	128 × 160	65K (16-bit)	2.0	x	x	x	x	x	-	-	-
6	L	3.9%	CSTN	128 × 160	65K (16-bit)	2.0	x	x	x	x	x	-	-	-
7	H	2.6%	TFT	170 × 220	262K (18-bit)	2.0	x	x	x	x	x	-	-	x
8	M	2.5%	TFT	240 × 320	262K (18-bit)	2.0	x	x	x	x	x	x	x	x
9	N	2.5%	TFT	128 × 160	65K (16-bit)	2.0	x	x	x	x	x	x	-	x
10	O	2.4%	TFT	176 × 220	262K (18-bit)	2.0	x	x	x	x	x	x	x	x

Yi, Maghoul and Pedersen (2008) studied the characteristics of mobile search queries submitted through several Yahoo! one-Search applications. In all, they worked with 40 million English language queries submitted by users in the US, Canada, Europe and Asia during a two month period in late

2007. The search application interfaces included an XHTML/WAP browser (http://m.yahoo.com), a java based interface (Yahoo! Go) and an SMS text messaging interface (Yahoo! Mobile SMS). Important characteristics of this study included the analysis of a multi-national, multi-interface data set of English language queries on an as yet unseen scale (20 million US and 20 million International queries).

Tables 10, 11 and 12. From Yi, Maghoul and Pedersen, 2008

Table 1: Query Distribution

		US	International
Total # of queries		20M	20M
# of unique queries		4.49M	3.7M
Avg. # of query repetition		4.46	5.41
# words per query			
All Queries	Avg	2.35	2.1
	Median	2	2
	StdDev	1.16	1.09
	Max	65	60
Uniq Queries	Avg	3.05	2.54
	Median	3	2
	StdDev	1.41	1.3
	Max	65	60
# characters per query			
All Queries	Avg	13.73	13.6
	Median	13	13
	StdDev	7.13	6.8
	Max	263	501
Uniq Queries	Avg	18.48	17.5
	Median	17	13
	StdDev	7.92	9.13
	Max	263	501

Table 2: US Mobile Query Categorization

	Unique Queries			All Queries		
Categories	% of queries	Avg. words per query	Avg. chars per query	% of queries	Avg. words per query	Avg. chars per query
Arts & Humanities	<1%	3.14	19.32	<1%	2.39	13.73
Automotive	2%	3.29	18.89	1%	2.60	14.48
Consumer Goods	2%	3.07	18.5	2%	2.28	13.70
Entertainment	44%	3.26	18.78	51%	2.55	14.68
Finance	1%	3.36	21.24	1%	2.18	12.39
Government & Politics	1%	3.05	20.99	<1%	2.87	17.52
Health & Pharma	2%	3.27	20.85	1%	2.57	16.36
Hobbies	<1%	3.06	19.04	<1%	2.49	15.67
International Interest	<1%	3.33	19.90	<1%	2.56	14.98
Life Stages	2%	3.33	21.15	1%	2.71	16.66
Miscellaneous	2%	3.17	18.71	2%	2.49	14.38
News	2%	3.21	19.35	2%	2.50	14.61
People	3%	2.73	17.18	5%	2.24	13.96
Reference	1%	3.64	21.89	<1%	2.75	16.91
Religion	1%	3.05	19.40	1%	2.17	14.33
Retail	5%	3.36	20.08	4%	2.35	14.21
Science	1%	3.13	19.70	1%	1.83	10.66
Small Business	2%	3.25	20.83	1%	2.57	16.22
Sports	3%	3.29	20.46	3%	2.40	14.23
Technology	6%	3.36	20.54	7%	2.19	12.74
Telecommunications	2%	3.49	21.05	2%	2.75	16.56
Travel	7%	3.34	20.03	7%	2.30	12.30
Uncategorized	12%	1.45	11.59	9%	1.26	8.98

Table 3: International Mobile Query Categorization

	Unique Queries			All Queries		
Categories	% of queries	Avg. words per query	Avg. chars per query	% of queries	Avg. words per query	Avg. chars per query
Arts & Humanities	<1%	2.94	18.26	<1%	2.94	14.66
Automotive	1%	3.01	17.49	1%	2.50	14.15
Consumer Goods	1%	2.81	17.13	1%	2.33	14.72
Entertainment	42%	2.88	18.30	47%	2.77	14.71
Finance	1%	2.95	18.51	1%	2.38	15.90
Government & Politics	<1%	2.55	18.86	<1%	2.55	15.77
Health & Pharma	1%	3.02	19.48	1%	2.48	13.62
Hobbies	<1%	2.80	18.21	<1%	2.80	15.56
International Interest	<1%	2.63	17.00	<1%	2.27	14.82
Life Stages	2%	2.80	18.81	1%	2.29	14.68
Miscellaneous	<1%	2.94	18.82	1%	2.93	15.36
News	1%	2.87	18.52	1%	2.87	14.81
People	3%	2.85	18.23	4%	2.85	14.24
Reference	<1%	3.48	21.78	<1%	3.48	17.53
Religion	<1%	2.43	17.97	<1%	2.43	15.04
Retail	3%	3.02	18.39	3%	2.21	15.93
Science	<1%	3.02	19.75	<1%	3.02	16.31
Small Business	1%	2.82	19.35	1%	2.40	14.53
Sports	2%	3.05	18.85	1%	2.35	14.99
Technology	5%	3.01	20.46	5%	2.11	15.97
Telecommunications	2%	3.13	21.48	2%	2.26	14.57
Travel	3%	2.30	17.75	2%	2.02	10.86
Uncategorized	28%	1.70	14.81	28%	1.45	13.63

Though personal entertainment ranked as the top category in both geographic areas, it is also clear that there are some regional differences. The US queries were more homogenous, longer queries with more words and a long tail of unique terms despite the similarity in category. Some variations among interfaces are seen and may be attributable to capabilities of devices. The authors conclude that "we believe mobile users are still figuring out ways they can utilize the new device and services, and their usage pattern is still evolving."

Another study by Church, Smyth, Bradley and Cotter (2008) looking at European mobile search patterns involved around 6 million search requests representing more than 260,000 unique mobile searchers. Data was collected over a 7 day period in 2006 and the authors compared study design characteristics with Kamvar and Baluja (2006, 2007) and Baeza-Yates et al. (2007) in the table below:

Table 13. From Church, Smyth, Bradley and Cotter, 2008

Key Parameters	Kamvar & Baluja (2005) [14]	Kamvar & Baluja (2007) [15]	Church et al. (2007) [6]	Baeza-Yates et al. (2007) [1]	Current Study
Coverage					
Regional	US	US	Europe	Japan	Europe
Search Engine	Google	Google	32 search engines	Yahoo!	32 search engines
Analysis Type					
Query (search input)	Y	Y	Y	Y	Y
Click-thru (Search output)	N	N	N	N	Y
Basic Statistics					
Number of Users	N/A	N/A	50,000	N/A	260,000
Number of Search Requests	1,000,000	1,000,000	420,000	N/A	6,000,000
Number of Unique Queries	N/A	N/A	91,000	1,000,000	600,000
Mean Terms per Query	2.3	2.7	2.1	2.3	2.2
Mean Chars per Query	15.5	16.8	13.0	7.9	13.8

Table 1: A comparison of summary statistics (approximate) for existing mobile search studies.

Important strengths of this approach included a click-thru analysis (surrogate for search success), large dataset and multiple search engines.

The Top 500 queries were classified according to whether they were informational (10.2%), navigational (29.4%) or transactional (60.4%) in nature. These data differ significantly in proportion from typical web search classification. For example, Jansen, Booth and Spink (2008) report percentages from their automatic classification of web queries as informational, navigational or transactional of 80.6%, 10.2% and 9.2% respectively.

They conducted an investigation into click-thru behavior in order to try to measure success (click-thru has been used as a crude surrogate for success). What they find is that for almost 90% of queries, no results are selected. Approximately 12% of Google queries are successful by this measure and among unique Google queries, approximately 24% lead to at least one click-thru. At the session level, about 41% result in selection of a search result. Of these, "35% of result selections lead to follow-on browsing with an average trail length of approximately 2.7." In sum, key differences are observed between unique searches, user searches, session searches and all searches suggesting that there is significant room for improvement. It may be that click-thru isn't always necessary to meet the information need of the user, browsing may be satisfying some of these needs with improved applications for mobile users.

The authors conclude that "the fact that almost 90% of searches failed to attract any result selections suggest that mobile search is very poor." They also describe mobile search as still in its infancy, that adult content still prevails and that mobile searching is analogous to desktop search in that short queries are used and the first few results are crucial. They also note that topics and taxonomies differ (adult content and transactional/navigational intent) and note that search engines are not tailoring interfaces to mobile users which results in poor link selection at the search result stage.

Table 14. From Church, Smyth, Bradley and Cotter, 2008

Category	% Top 500 Queries [6] (Church et al 2007)	% Top 500 Queries (Current Study)
Adult	53.5	61.2
Email, Messaging & Chat	8.4	9.1
Search & Finding Things	8.4	7.2
Entertainment	8.1	5.9
Multimedia	10.4	5
Socializing/Dating	1.8	2.2
Sport	1.1	2.2
Shopping & eCommerce	1.6	2.1
Games	2.5	1.5
User Generated Content (UGC)	0	1.1
Unknown/Unclassified	1.9	0.7
Mobile Applications, Websites & Technologies	1.4	0.7
Information	0.3	0.6
Auto	0.3	0.2
News/Weather	0.3	0.1
Employment	0.1	0.1
Local Services	0.3	0.1

Table 3: A comparison of the top 500 queries (averaged over 7-days) classified by query category for the current study and the Church et al. study.

On the topic of interaction style and the impact this has on search among mobile device users, the authors indicate "It is interesting to note that despite the text-input challenges presented by mobile devices, mobile searchers do appear to submit similar length queries to those used in Web search, at least during the early years of Web search when average query lengths were reported to be in the region of 2.3 terms." This is followed by the note that "of course the arrival of next-generation touch-based displays offers a whole new set of interaction modalities."

Concerned about the amount of time it takes a typical mobile phone user to enter in a set of query terms, Kamvar and Baluja (2008) conducted an experiment examining the effect of query suggestion on mobile users. Users of Motorola RAZR phones were recruited and each user was assigned one of six different interfaces providing query suggestions. Users were instructed to enter predefined query topics and avail themselves of the query suggestion system. NASA task load index was used to measure workload and information on their query habits was recorded.

Of the users who were shown suggestions, 100% accepted at least one suggestion. For the most part, suggestions appeared to be accepted quickly. The authors observed that it was hard for users to make the cost-benefit analysis of time saved the keystroking process while entering a query versus accepting a query suggestion. Fewer suggestions seemed to improve the odds of a user selecting one and the movement of suggestions in the list hindered acceptance. Comparing these findings with other devices and desktop systems was recommended.

Wishing to investigate search pattern differences among devices, Kamvar, Patel and Yu (2009) conducted a web transaction log analysis of the search patterns of desktop, iPhone and conventional mobile phone users. During a 35-day period in the summer of 2008, a random subset of 100,000 queries representing 10,000 users were collected for each interface. The data were limited to search users submitting English language queries. The following table summarizes results across the three devices:

Table 15. From Kamvar, Patel and Yu, 2009

Table 6: Single-Session User Statistics

	Computer	iPhone	Mobile
percent of users who engaged in one search session over the 35-day period	29.4[6]	22.89	42.6
average number of queries per search sessions	1.88	1.89	1.74
average characters per query	18.00	16.04	15.86
average words per query	2.795	2.589	2.489

This was an extensive and controlled comparison of search users which demonstrated that search usage is more focused for the average mobile user than average desktop user. They found that search on high end phones resembled that of desktop use and that query length on the iPhone was similar to that of the desktop. Desktop and iPhone search query diversity appeared to be similar as did the frequency of unique queries. Mobile phone query length was much shorter and queries were much less diverse. It was assumed that iPhone users were more likely to use tailored applications for contextual content (rather than Google search) in part because iPhone and desktop contextual searching was similar. Significantly less than on the mobile phone, iPhone adult content searches were more similar to desktop searching. Interestingly, the diversity of information needs per user was greatest among the iPhone users. Desktop users still showed the highest number of queries per session per user followed by iPhone users then mobile phone users. Frequency of search followed the same pattern being highest among desktop users, then iPhone and mobile phone. From this, the authors concluded that mobile search is still a secondary mode of searching and make the following important recommendation:

> "We suggest that for the higher-end phones, a close integration computer-based interface (in terms with the of personalization standard and available feature set) would be beneficial for the user, since these phones seem to be treated as an extension of the users' computer."

In an effort to better understand mobile user intent, Church and Smyth (2009) conducted a four-week diary study of mobile information needs. Their focus was on topics of interest and the impact of contextual factors like location and time. Significant findings of the study included a modification of the three classifications of mobile search from the traditional web search model of transactional, navigational and informational to informational, geographical and personal information management (PIM). The distribution by diary entry and a comparison between mobile and non-mobile are indicated in the tables below.

Tables 16 and 17. From Church and Smyth, 2009

Goal	% Entries
1. Informational	58.3
2. Geographical	31.1
2.1 Local Explicit	12.8
2.2 Local Implicit	14.8
2.3 Directions	3.5
3. Personal Information Management (PIM)	10.6

Table 1. Results of classifying diary entries by intent.

Goal	Mobile	Non Mobile
1. Informational	64%	36%
2. Geographical	75%	25%
3. Personal Information Management (PIM)	65%	35%

Table 3. Percentage of diary entries associated with each goal/intent (i.e. informational, geographical and PIM) by the location context, i.e. mobile and non-mobile. Non-mobile refers to entries generated while the user is at home, at work or in college while mobile refers to entries generated in all other instances, e.g. commuting, traveling, etc.

Another interesting finding was that classification of diary entries by topic looked much different from that of web search. Local services, travel and commuting and general information were more often indicated than entertainment. The table below provides a complete list.

Table 18. From Church and Smyth, 2009

Topic	% Entries	% Users
Local Services	24.2	95
Travel & Commuting	20.2	85
General Information	15.6	85
Entertainment	12.8	75
Trivia	6.4	45
Sport	3.5	30
Email & Social Networking	3.2	40
General Shopping	3.0	25
Cooking, recipes, ingredients	2.2	35
To do/schedule	2.0	20
Stocks/finance	1.7	30
News/Weather	1.5	20
Misc	1.2	20
Personal Info	1.2	25
Education	0.5	5
Employment	0.5	5
Auto	0.5	5

Table 2. Results of classifying diary entries by topics.

Context

Context really does appear to be King in mobile search. But context has many facets and an equal number of considerations must be taken into account when developing for the mobile context. Dey and Abowd (1999) define context as follows:

> "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."

This may include the computing environment, user environment and location and the physical environment. From lighting and noise level to network connection, interaction devices, and social situation, context can vary significantly. The anywhere, anytime nature of mobile computing really appeals to us, despite the fact that most of us follow a very similar path of activity on a daily basis (Bayir, Demirbas and Eagle, 2009).

Dey and Abowd also claim that some elements of context are more important than others. Location, identity, activity and time are the primary components of context and can be expressed as where you are, who you are with, and what objects are around you. Context aware applications are ones that use "context to provide relevant information and/or services to the user, where relevancy depends on the user's task."

In their work on mobile search intent, Church and Smyth (2009) uncovered just how significant a factor context is in mobile search. Reshaping the long accepted classification of web search from queries of informational, transactional or navigational intent to queries of information, geographical or PIM intent, mobile search is indeed different from web search. Supporting the frequent call for search interfaces tailored to mobile use, Church and Smyth (2009) focus on contextual cues beyond location and time to activity and social interaction/conversation.

Improving mobile search

Jones, Buchanan and Thimbleby (2002) make a critical contribution to the improvement of mobile search focusing on search failures and how to improve them. Comparing a WAP browser to a PDA style interface and the Google 'classic' interface, one of the most significant differences in search using a small screen is a limited result set. Users took almost twice as long on average to perform the same searching task using the WAP browser and tended to be less successful. For the PDA interface, search was more similar to that of the 'classic' interface and differences were found to be insignificant. When users failed to complete a search task using all three interfaces, they failed badly taking 2-3 times the amount of time as successful searches.

Bila, Ronda, Mohomed, Truong & de Lara (2007) investigate the value of increased customizability to improve search interaction on mobile devices. They use a technique called Reusable End-User Customization (REUC) developed in a software tool format, PageTailor, to store a user's page preferences for a given web site and apply those parameters whenever that page is presented. Execution time on the PDA is compared with that of a desktop system to illustrate where improvements might be sought. The big advantages to this are improvements to readability and usability of favorite or frequently accessed sites over time. Disadvantages include significant time investments to 'tailor' the page the first time and problems with links and underlying code in the restructuring process.

List based search interfaces are compared with (Conceptual Reorganization of Documents) CREDO by Carpineto, Mizzaro, Romano and Snidero (2009). Their findings on the use of Credino and SmartCREDO, for PDAs and cell phones (respectively) suggest that for subtopic queries, clustering search results can be at least as effective as more traditional search engines. In addition, though their findings suggested that the smaller the device, the poorer the search retrieval, mobile search can be facilitated when retrieval clustering is employed.

In a pair of papers with some shared authorship, the notion of focusing on answers rather than questions in the search process on mobile devices is explored. To test this notion, Jones, Buchanan, Harper and Xech (2007) provided a list of queries specific to a users' location and presented this to a test group. These 'in situ' queries were found to be positively influential with the testing group. In the 'companion' paper, Arter et al., 2007 evaluated their prototype application *QnotA,* based on the answers not questions approach. The tool provides the user an alternative to the traditional search/browsing experience and provides location specific searches performed by other users. While some searches were considered too generic to gain much interest from the test group, many searches were deemed specific and interesting enough to be useful. Though subjects were 'local' to the areas of interest, most reported learning something from the queries.

Church and Smyth (2008) aptly state that "Limited screen- space, restricted text-input and interactivity, and impatient users all conspire to exacerbate the shortcomings of modern Web search." In an effort to overcome this, the authors proposed a prototype search interface aimed at combining location, time, and community preferences to improve mobile search. Query and result selections for a specific geographic area were mapped with yellow and red balloons, respectively. In addition, the users could access either of two slider bars at the bottom of the screen adjusting for temporality (earlier > now), and query similarity (broad > narrow). This allows the user to constrain the 'community' focus of the search.

6. METHODS

General

There are so many possible approaches to take in a research project that narrowing in on a proper methodology is difficult. Should it be conducted in a laboratory, should it be more naturalistic? Will the trials be experimental or observational in nature? Will some type of log analysis be used to improve data capture? What elements should be randomized, what are the dependent variables? Should surveys be included? What data will be collected? How and in what form/format? What is the best analytical approach for the data?

Early HCI work (see Figure below from Marchionini and Siebert, 1991) resulted in some persistent and robust models for measuring, testing and even predicting interaction. These models are still employed today in the development of new tools and can provide important insight into the adoption and continued use of them. As we stretch out beyond the asymptote of adoption with computing technologies, increasingly our focus becomes that of the experienced user and of use of devices that takes on new dimensions. The advent of multi-touch interaction on mobile devices is a pivotal shift, and coupled with location based services represents interaction in a new dimension. Some might argue that using existing methods to evaluate new interaction modalities may be insufficient or inappropriate. It remains to be seen, however, how these models compare across devices with differing interaction styles.

	Science		**Development Technique (engineering)**
	Descriptive	Predictive	
Human Characteristics	Norman 4-stages Norman slips (Perceptual Models) (Cognitive Psychology) Egan et al user abilities	Routine cognitive skill GOMS, NGOMSL, CCT Unified cognitive models MHP, Soar, ACT* ETIT Fitts' Law	KL-Model PUMS Fitts' Law
Machine	Input device taxonomies Buxton et al; Foley et al; Mackinlay et al.	State-transition (CSP, CCS, ESTEREL), Squeak, PIE, Alexander PAC, PPS Grammar Models BNF, CLG, TAG, ETAG, DTAG Fitts' Law	Fitts' Law

Figure 2. Selected HCI Models

Figure 9. From Marchionini and Siebert, 1991

All of these questions and more go into the planning and execution of a good study. For studies involving multiple dimensions (e.g. display size, interaction style and search) the possibilities are truly overwhelming. However, for the sake of brevity, only the main methods being explored for each area will be expounded on here. For display size, studies most often involve a controlled laboratory setting where users are observed and/or videotaped and, in some cases, their eyes/movements are tracked to provide for a more detailed analysis. In evaluating interaction style, many methods are employed some stemming from engineering models like Fitts's Law, the Hick Hyman Law, or Steering Law. Others originate from psychology looking at cognitive models of interaction such as cognitive walkthrough, heuristic evaluation, ACT-R, GOMS, KLM and so on. In search studies, many different methods have been employed over the years from surveys and structured interviews to laboratory experiments and, in recent years, web transaction log analysis.

Exploratory, Descriptive or Explanatory

Kelly (2009) describes studies in IIR as falling into in three main categories: exploratory, descriptive and explanatory. Sometimes the relationship between these types of research appears to be chronological and at other times, issue focused (e.g. the nature of search). Increasingly, researchers are taking hybrid approaches to understanding different aspects of user behavior trying to leverage the strengths of each approach. In no way unique to IIR, these three areas represent most of the possible research approaches in ILS.

Observational or Experimental, Field Test or Laboratory Test

Sometimes observational studies are considered more valuable because they make little prescription for what the user's intentions are by merely observing the user in his/her own domain on his/her own terms (Blackwell, 2010). This approach is often taken when influencing the user will result in an undesirable outcome. A good example might be open or exploratory search where the objective is to simply allow the user to perform a search without any specific direction. Studies of search intent and/or search success might be well suited to this approach. A major drawback to this approach is that there are many unknowns which may make the findings hard to generalize. In the realm of mobile computing, observational studies are preferred but extremely difficult to execute successfully. So many variables are in play when trying to record user behavior in a mobile context, particularly the influences of the external environment. For this reason and others, log analysis has been a preferred route for large scale mobile device research.

Controlled studies conducted in a laboratory environment with specific protocols and enrollment criteria are typically described as experimental studies. Increasingly, information and library science (ILS) research includes analysis of an experimental nature. HCI studies almost always involve controlled experiments in a laboratory environment. Without control, the fine details being monitored would likely not be sampled well enough to be widely applicable and/or generalizable and thus, wouldn't offer much more than anecdotal information. An area where the field is fairly mature in terms of study design, HCI studies typically strive to relate findings to the prior literature using well established methods with many protocol details which need to be adhered to.

Qualitative, quantitative and hybrid approaches

While the discussion about whether to take a quantitative approach or a qualitative approach to any given research question can be intrinsic to the question at hand (and hotly debated still), more and more often we see hybrid approaches providing a wealth of insight. In attempting to model human behavior and human cognitive processes in the use of information technology, it has been enormously helpful and productive to produce and refine quantitative measures of behavior/interaction. However, these models are limited and do not provide a holistic view of user experience. Hybrid approaches, combining multiple methods of data collection, often produce a more meaningful and comprehensive understanding of human experience.

To narrow the scope of possible methodologies, an outline is included in the table below. In the left hand column is a [typical] Interactive Information Retrieval Research (IIR) Scenario from Kelly (2009). In the right hand column, is a proposed scenario to be outlined in more detail in the following sections.

Table 19. A [typical] Interactive Information Retrieval Research (IIR) Scenario from Kelly (2009) in the left hand column and a proposed scenario in the right hand column

A [typical] IIR Research Scenario	Proposed Research Scenario
A researcher has developed two experimental IR systems and would like to test them against one another and a baseline. These three systems will be called System A (the baseline), System B and System C (note that *system type* functions as one variable, with three levels). Subjects are given six search tasks to complete which ask them to find documents that are relevant to pre-determined topics. Each subject completes two searches on each system (a within subjects design). The researcher is interested in comparing the three systems using the measures listed below. These measures are organized according to the instrument used to collect the data. **Demographic Questionnaire** *Sex* of Subject [Male or Female] **Pre-Task Questionnaire** How *familiar* are you with this topic? [5-point scale, where 1= know nothing about the topic and 5=know details] **Post-System Questionnaire** *Usability* [5-point scale, where 1=strongly disagree and 5=strongly agree]: It was easy to find relevant documents with the system. **Exit questionnaire** *Preference*: Which system did you prefer? [System A, B or C] **System Logs** *Performance* • Average session-based nDCG *Interaction* • Number of queries issued • Query length Fig. 11.1 IIR Research Scenario.	A researcher is comparing three distinct computing environments: Desktop System (baseline), Tablet/Slate/Netbook and Mobile Device/Phone. Input and Output interaction methods will be observed through the modeling of low level tasks (typing, pointing, scrolling and pinch/expand) and searching tasks (simple search, faceted search and exploratory search). **Demographic Questionnaire** • Age, gender, handedness, prior use of mobile devices, other prior experience **Pre-Task Questionnaire/Semi-structured interview** • Understand current usage practices, familiarity with multi-touch and mobile device preferences **Post-Task Questionnaire** • NASA TLX to be administered after each task on each device **Post-System Questionnaire** • UTAUT to be used for overall impression of the technology • MPUQ to be used for mobile device usability analysis **Video and Audiotaping, and screen capture** • Fitts's Law and KLM Timing data will be both modeled and extrapolated from video files

Surveys/Questionnaires

Survey instruments abound in ILS and particularly in the realm of HCI. Two key components of a good instrument, reliability and validity, take time and repeated use of the same instrument to measure and understand (Kelly, 2009). For this reason and to improve the ability to relate the findings in one study to others with similar methodology, the use of well tested survey questions is critical. Two such survey instruments are outlined below.

<u>Acceptance</u>

The Unified Theory of Acceptance and Use of Technology, which synthesizes several technology acceptance models (including The Theory of Reasoned Action (TRA), The Theory of Planned Behavior (TPB), and The Technology Acceptance Model (TAM)), (Venkatesh et al., 2003) has been tested and validated even cross culturally (Oshylansky, Cairns and Thimbleby, 2008). The main constructs of UTAUT which include performance expectancy, effort expectancy, social influence, and facilitating conditions, are affected by age, gender, experience and voluntariness of use and are considered direct determinants of use intention and behavior. UTAUT is perhaps one of the most widely accepted and used theories behind surveying IS acceptance and use behavior.

<u>Usability</u>

A usability questionnaire designed for mobile devices, the MPUQ (mobile phone usability questionnaire) was developed by Ryu and Smith-Jackson (2005). This work was based on the definition of usability offered by ISO 9241-11 (see definitions) and a synthesis of several usability questionnaires including the Questionnaire for User Interface Satisfaction (QUIS) and the Software Usability Measurement Inventory (SUMI). In a psychometric analysis of the questions, Ryu and Smith-Jackson (2006) identified 72 questions in six main categories: (1) Ease of learning and use, (2) Assistance with operation and problem solving, (3) Emotional aspect and multimedia capabilities, (4) Commands and minimal memory load, (5) Efficiency and control, and (6) Typical tasks for mobile phones (TTMP). Combining elements of the MPUQ and UTAUT to survey subjects on their qualitative impression of the devices planned for use in this research.

Interviews (structured, semi-structured or unstructured)

Structured interviews, praised for their generalizability, are often unusable in a preliminary analysis of a research question. Unstructured interviews, while extremely helpful in the early stages of evaluating a novel research question are hard to generalize from. Semi-structured interviews offer an approach that provide structure and allow for probing or tailoring to the subject's situation or response(s) (Wildemuth, 2009).

We will conduct semi-structured interviews as a component of the study. This will help us gather demographic information about the user, determine the degree of familiarity with computing systems (including mobile systems) and help us chart a typical use pattern among subjects. Structure will provide homogeneity across some elements of the interview. By allowing the interviewer an opportunity to present questions based on a subject's response, the subject's responses may be probed further (e.g. if a subjects uses their mobile device for internet searching, the interviewer can find out more about frequency and type of searching) and certain questions may be omitted (e.g. if a subject does not use their mobile device for internet searching, no further questions in this topic area will be asked).

Usability and Task Analysis

It may be a known phenomenon—that usability testing is insufficiently conducted for most web sites and most mobile devices. From interaction style to widgets, not enough work on usability is being done before hardware and software solutions hit the market. Microsoft, for example, has had the reputation that it uses the general public as a beta test environment. In the mobile device arena, new input techniques have been known to be released without much if any prior usability testing outside of the design process–" all too common are published reports stating only qualitative results steeped in anecdote, or, worse yet, testimonials unsupported by empirical data (Mackenzie and Soukoureff, 2002)." Many researchers have provided robust and easy to implement tools for usability evaluations which provide significant benefit for designers, manufacturers and users alike. Molich and Nielsen (1990) developed nine usability heuristics: use a simple and natural dialogue, speak the user's language, minimize user memory load, be consistent, provide feedback, provide clearly marked exits, provide shortcuts, provide good error messages, and prevent errors. Seemingly self-evident, these heuristics have transformed the way designers approach their work and remain in regular use today.

Favored because it is practical and liked for its human centered focus, task analysis has helped transform 'help' systems from detailed technical manuals to user-friendly, interactive and task-based online tools. Covering the full range of activities associated with a given task including mental activities, degree of complexity and environmental conditions, task analysis has informed hardware design, required training, procedural design (workflow) and can help identify areas where automation may be employed. Though this research will involve task analysis in general, our focus will be on the execution time to complete a task. In formulating the tasks, we will aim to provide a representative sample of tasks for the most common types of interaction a typical user might engage in. In addition, we will constrain the tasks so that cognitive load (mental thinking time) might be kept to a minimum in order to better compare across devices. Our aim is to also include an open-ended type of task which would allow the subject to inform task selection and execution, an important consideration in light of the substantial shift in interaction modalities among mobile devices.

Workload

The NASA Task Load Index (TLX), a subjective workload assessment tool, provides a weighted measure of workload based on six subscales: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration. Used widely in HCI for discrete task based encounters with human-machine systems, the NASA TLX is a quick and robust measure at the discrete task level that can provide a sensitive and reliable estimate of workload across task types. This measure will be obtained for each user after each task. This will allow comparisons in perceived workload of the same task among differing device display sizes and interaction styles.

Think/Talk Aloud

Exploratory search, that is a search task that is not necessarily prescribed for the subject, can provide rich information about user experience for more complex tasks. Though the think aloud protocol may be cognitively complex for users, there is some evidence that providing a period of training and redundant methods of data collection (e.g. a microphone for recording the subject's voice) when using this approach may improve the data collection process (Kelly, 2009). In addition to this, a variant, the talk aloud approach, may yield informative results without the same cognitive overhead. While we expect to see and measure differences in interaction between devices with differing interaction styles and display sizes, we expect that there are many other differences that may be more observable with more open ended search tasks and a less quantitatively focused data collection method. Some combination of the think aloud and talk aloud approaches may be warranted for this phase of testing.

Scientific Models in HCI

Outside of usability testing, assessment methods include standard user reports and empirical testing, analytical observational methods and cognitive walkthroughs. While many different approaches to "assess user performance prior to usability testing" are taken, predicting performance requires a cognitive model with generative qualities and explanatory power (Cox et al., 2008). A variety of theoretical models have been used to measure and predict aspects of usability in HCI. Fitts's Law (and Shannon's Formulation), the Hick Hyman Law and Steering Law, each modeling central modes of human movement, played a pivotal role in early interaction device development. A great deal of work has been done using Fitts's reciprocal tapping task (e.g. Mackenzie, 1992) to evaluate input devices.

Fitts's Law

A robust model of human movement involving the entire receptor-neural-effector system, Fitts's Law describes the relationship between speed and accuracy and the inherent tradeoff made between the two to transmit information required to organize motor behavior. In his 1954 work "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement," Fitts used information theory to examine and relate the basic concepts: amount of information, noise, channel capacity, and rate of information transmission. While holding all relevant stimulus conditions constant (outside the subject), subjects were asked to make rapid, uniform and automatic/overlearned responses. This allowed for the development of a model to predict the time required to move to a target as a function of the size of the target and the distance to it. In sum, "the model predicts movement time as a function of a task's index of difficulty—the logarithm of the ratio of target amplitude to target width" (Mackenzie, 1992 p. 97).

Shannon's formulation of Fitts's law (Mackenzie, 1992) in common use today is:

$$T = a + b\log_2\left(1 + \frac{D}{W}\right)$$

Where:
T= the average time it takes to move to the target
a= the start/stop time of a device
b= the speed of the a (a constant)
D= distance from the starting point to the center of a target
W= width of a target measured along the axis of motion (allowed error tolerance of the target)

An illustration of the reciprocal tapping task used in the original Fitts's studies is included below:

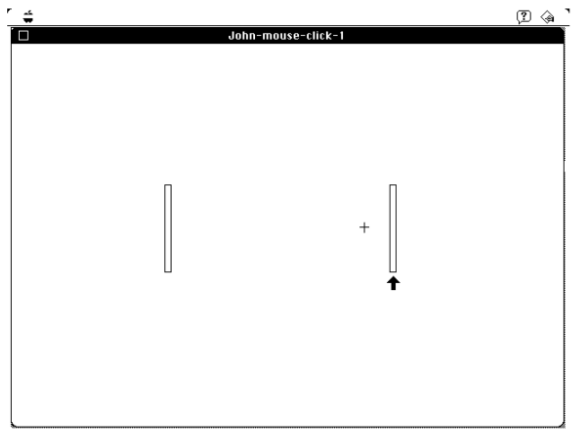

Figure 1. A typical screen showing Fitts' reciprocal tapping task (point-and-select) as implemented for a CRT display and any input device.

Figure 10. From Mackenzie & Buxton, 1993

The starting point for Fitts' law in HCI research was in the work of Card et al. (1978) and it is integral to the Model Human Processor concept developed by Card, Moran and Newell (1983) (Mackenzie, 1992). Card et al. (1978), investigating the performance of various input devices used the index of difficulty or (ID) from Fitts's law to provide good performance estimates for pointing tasks using the joystick and mouse. The equation for ID is as follows:

$$ID = \log_2\left(\frac{D}{W} + 1\right).$$

Where:
ID=index of difficulty
D=distance to the center of a target
W=width of a target measured along the axis of motion (allowed error tolerance of the target)

In the linear equation that results: ID=$a + b \log_2 (D/W + 1)$, where a and b are constants given by the regression, the reciprocal of b is index of performance (IP), a measure that can be used to compare "human rate of information processing for the movement task" (MacKenzie and Buxton, 1993). Fitt's Law has proven to be a robust, well studied model and is applied regularly to myriad usability issues. Recently, Fitts's Law has been evaluated for predicting task execution times on mobile devices. St. Amant, Horton and Ritter used a variety of predictive models to investigate

Goals, Operators, Methods and Selection Rules (GOMS)

In an effort to improve our ability to understand, observe, and make predictions about human-computer interaction(s,) Card, Moran and Newell developed a human information processor model they labeled the Model Human Processor (MHP) in their book "The Psychology of Human Computer Interaction" in 1983. As originally laid out, the system consists of perceptual, cognitive and motor subsystems to perform the necessary steps to sense, process, retrieve, store and respond to a given set of instructions for a task. According to the MHP, users will attempt to solve a problem using the most efficient means allowed by the system (task environment and human abilities) according to two key principles of the Model Human Processor: the Rationality Principle and the Problem Space Principle. The system is composed of 'empirical' timings to complete these steps for any given task. If a task is modeled down to each of its constituent elements, predictions can be made using the model for any given task. But, GOMS modeling is tedious and often

Their specialized version of the MHP that can be applied to human computer interaction the Goals, Operators, Methods and Selection rules (GOMS) model has been widely considered a fundamental tool for observing actions in the human-computer system. It has been a valuable tool with which to aid system developer(s) in predicting acceptability of a system using model-based predicted interaction times. Since 1983, many researchers have tested, evaluated, and made modifications to the original CMN-GOMS concept. As a result, several models have been proposed over the years and are outlined in the table below:

Table 20. Variants of the Goals, Operators, Methods and Selection rules (GOMS) model

GOMS or Goals, Operators, Methods and Selection Rules also known as **CMN-GOMS** or Card, Moran and Newell GOMS	The original Card, Moran and Newell theory based on the Model Human Processor (Card, Moran and Newell, 1983)	Establishes the concept that user interaction with a computing system consists of the following: • **G**oals: the intent of the user • **O**perators: elementary perceptual, motor or cognitive actions taken by the user in interacting with the system to accomplish the goal • **M**ethods: an internalized set of procedures or algorithm that the user employs to accomplish the goal • **S**election rules: A set of context specific rules the user employs to determine how to complete a given component of the goal task Tasks must be serial in order to be modeled using CMN-GOMS. CMN-GOMS provides quantitative predictions of task execution in repeatable program form without reference to a specific method.
KLM or Keystroke Level Model	Simplified GOMS, the (Card, Moran and Newell, 1980)	An original, simplified version of CMN-GOMS which seeks to simplify the inherent perceptual and cognitive functions in GOMS while still providing a method for predicting user performance. The basic operators outlined by Card, Moran and Newell are: • K: keyboarding • P: pointing (as with a mouse) • H: homing • D: drawing • M: mental thinking time • R: system response time Tasks must be serial in order to be modeled using the KLM. The KLM provides quantitative predictions of task execution without given a specific method.
NGOMSL or Natural GOMS language	Structured natural language notation for representing GOMS models. (Kieras, 1996)	This method maps in a one-to-one relationship to cognitive complexity theory methods of cognitive architecture. In addition to providing quantitative information that can also be provided by CMN-GOMS and the KLM, this method has been shown to be valuable in predicting procedural learning time. Tasks must be serial in order to be modeled using NGOMSL.

CPM-GOMS or Cognitive-Perceptual Motor GOMS	Requires a specific level of analysis where the acts are cognitive, perceptual or motor modeled on the PERT critical path to complete a task. (John and Kieras, 1996)	This method allows for modeling of parallel processes. These models assume extreme expertise in the user and begin with the CMN-GOMS model of the task. The extreme expertise assumption means that with task that involve variability in say, display position, then visual search would have to be included in the prediction, it also limits the degree of method selection since it assumes that method choice is already clear.
QGOMS or Quick GOMS	Also called "quick and dirty" GOMS or "back of the envelope" GOMS, this approach is intended to produce quick estimates using advanced features (Beard, Smith and Denelsbeck, 1996)	This variation on the original GOMS method was developed by Beard, Smith and Denelsbeck (1996) and uses a reduced number of modeling constructs, highly modified selection rule mechanisms, and a graphical tree structure and direct manipulation tool for visualization and manipulation.

Because there can be significant differences based on research goals in which GOMS method to use, John and Kieras, (1996) developed a set of criteria to help select which model to use. First determine:

1. the degree of goal-directedness in the user's purposes for using the system,
2. the degree of routinized skill involved in the user's task,
3. the degree to which the interaction is under the control of the user versus the computer system or other agents involved in the task, and
4. the sequentiality of the user's task.

The chart below is a helpful outline for matching task type to method depending upon whether the model is investigating sequential or parallel actions within the system.

Task Type / Design Information	Goal-Directed, Routine Cognitive Skill with Passive or Active Systems	
	Sequential	Parallel
Functionality: Coverage	Any GOMS	Any GOMS
Functionality: Consistency	NGOMSL	
Operator Sequence	CMN-GOMS NGOMSL	CPM-GOMS (see text)
Execution Time	KLM CMN-GOMS NGOMSL	CPM-GOMS
Procedure Learning Time	NGOMSL	
Error Recovery	Any GOMS	Any GOMS

Figure 1. GOMS techniques available for different combinations of task type and the type of design information desired. Note that only tasks that are goal-directed, routine cognitive skills are included, and information types not provided by GOMS models are not shown.

Figure 11. From John and Kieras, 1996, Which GOMS?, p.8

While GOMS is a preferred and well used model for usability in HCI research, there are many reasons for and against its use. The table below outlines some of the pros and cons of GOMS:

Table 21. GOMS Pros and Cons

GOMS Pros and Cons	
Pros	Cons
• Well established for use in HCI research • One of the most mature models of human performance • Predictive • Descriptive • Representative of how a user performs tasks with a given system • Provides explanatory basis for time estimates • Prescriptive • Instructive: once a model has been produced for a certain task, it can be used to train new users	• Requires 'expert' users • Model does not account for errors • While elementary, perceptual and motor components of skilled behavior are explained well by the model, cognitive processes are accounted for but less well explained by the model • Mental workload is unaddressed • Fatigue is unaddressed/unrecognized • Differences among users is unaddressed • Social and organizational impact of the system is unaddressed • Acceptability and enjoyment are unaddressed

<u>KLM</u>

The simplest version of GOMS, the Keystroke Level Model is a method for modeling, predicting, evaluating and validating the execution of a user-computer system task. The Keystroke-Level Model takes a systematic approach to determining the time it takes for any individual to interact with a computer system accounting for system response time and mental thinking time. The model defines a series of operators:

K: keyboarding
P: pointing (as with a mouse)
H: homing
D: drawing
M: mental thinking time
R: system response time

Taking a measurement of user typing speed, Card, Moran and Newell were able to create a model wherein execution times could be predicted based on the KLM model assigned to a given task. This concept has been validated in a number of follow up studies and though KLM has its pitfalls, particularly in that it is hard to apply in a naturalistic environment and is predicated on an expert level of performance without error, it is extremely robust in its predictions and good for use in comparing different devices.

Because the current ubiquitous Windows GUI has only a few operators per task: mouse pointing and clicking and keyboard press, the low level Keystroke Level Model is a reasonable choice for looking closely at task execution times using this system (and combination of input devices) (Bonto-Kane, 2006). The table below indicates average execution times for each KLM operator in the Windows GUI environment. Bonto-Kane also points out that KLM prediction models should be empirically identified, tested and validated. This is even more important in an era where new interaction styles like multi-touch are emerging.

Table 22. From Bonto-Kane, 2006 citing an online Kieras source

Operators	Average time
K – Keystroke (0.12s to 1.2s for ordinary user) Pressing a key or button on the keyboard. Different experience levels have different times. Pressing SHIFT or CONTROL key is a separate keystroke. Use type operator T(n) for a series of n Ks done as a unit.	0.28s
P – Point (0.8s to 1.5s for ordinary user) Point with mouse to a target on display. Generally follows Fitts' Law (use this if possible).	1.1s
B – Press/release mouse button (0.1s; click is 0.2s) Highly practiced, simple reaction.	0.3s
H – Home hands to keyboard or mouse	0.4s
W – Wait for system response Applicable only when the user is idle because he cannot continue until after system response. An estimate is done from system behavior Wait time is often essentially zero in modern systems.	
M – Mental act of thinking Represents pauses for routine activity (not-problem solving) New users must often pause to remember or verify every step. Expereienced users pause and think only when logically necessary. Estimates range from 0.6s to 1.35s	1.2s

Table 1: Keystroke operators and their average execution times (Source, Keiras online source)

Early modeling and testing with GOMS and KLM produced some dramatic results. Predictive models afforded a quantitative measure of the potential pitfalls, savings or efficiencies of new information systems (Gray, John, and Atwood, 1992). Bringing together the work of Kieras (2001) on demonstrating how KLM can be used to model (and improve) execution times and Hemminger's (1992) recommendation that the only response time that is acceptable is near zero, it is apparent that cognitive complexity can be significantly reduced by reducing execution times. This is a critical benefit that KLM modeling affords. Because the model operates at such a low level of interaction the gains per action can be enormous (Bonto-Kane, 2006).

KLM in the mobile context

Bonto-Kane (2006) highlights the gap in research in identifying empirical time data for KLM operators on mobile devices. Because these data are device and interaction style dependent, such information will need to be obtained and verified empirically for mobile devices just as it has been for the desktop. CogTool has been modified to include prediction times for a variety of mobile devices (Luo and John, 2005 and John and Suzuki, 2009).

Sears and Zha (2003) carried out a study comparing the use of three different sizes of soft keyboards (small, medium and large). Though they found no significant differences between the three keyboards, they did propose some refinements to the KLM model based on their observations. Because the trial involved transitioning between a primary and secondary keyboard (abc, 123), they proposed the following modifications to the operators:

T = total time to complete the task
t_1 = time for the first key press when beginning a new task
t_d = time to make a decision that a transition is required
t_r = time to recover from a transition and complete the subsequent key pres
t_k = time for each additional keystroke (not addressed by t_1, t_d, or t_r)
c = number of characters required by the task
c_s = number of shifted characters (e.g.,uppercaselettersoralternativesymbols)
c_t = number of transitions between keyboards required by the task

The change this represents to the total task completion time is then represented as follows:

$$T = t_1 + [(t_d + t_r)c_t] + [t_k(c + c_s - c_t - 1)]$$

This adjustment accounts for the fact that it takes more time (and mental load) to initiate the task, to transition to another keyboard and to work with symbols or uppercase letters.

Myung (2004) used KLM to model Korean language text entry using mobile phone keypads. Because the phone used only allowed keypresses for interaction, the model was simplified to only include the times for keystroking and for mental thinking time. First, a validation of the model was conducted and a refinement of the placement of mental operators was implemented. Because the Korean language has not fully adopted a keyboard layout, Myung used this opportunity to test a proposed layout that was intended to reduce cognitive distance by placing consonants in alphabetical order. Using the predictive model, Myung was able to evaluate several possible layouts and determine which of these could be expected to perform the best.

Buranatrived and Vickers (2004) used KLM to predict task execution time on a PDA and compare it to that on a WAP phone. Their findings are summarized in the table below:

Table 23. From Buranatrived and Vickers, 2004

Table 5. Theoretical vs. actual task times

	Stock broking		Ticket purchasing	
	Phone	PDA	Phone	PDA
Theoretical duration (sec.)	16	19	199	235
Mean actual duration (sec.)	20	18	217	226

Their findings suggest that the model more accurately predicts task execution time on a PDA than on a WAP phone.

Luo and John (2005) tested the use of CogTool to provide predicted times for interaction using a PalmOS device. A typical model of interaction produced by CogTool might look something like this:

```
(klm-p (klm-goal klm
(think)
(look-at "Museums")
(press-button "Museums")
(think)
(look-at "-graffiti-")
(press-button "-graffiti-")
(think)
(press-button)
(think)
(look-at "MET")
(press-button "MET")
… …
```

Figure 2 Example KLM code for Method 3

Figure 12. From Luo and John, 2005

Predicted times were then compared to actual empirical timings generated using EventLogger on the PalmOS device. Their findings suggested that the model was applicable to these interactions within a 6% margin of error. The highest prediction error was associated with stylus (gestural) use and attributable to users with no prior experience with a stylus.

Holleis et al. (2007) make important strides in trying to map some new interactions using the KLM model. Finding the original operators insufficient in explaining the full range of possible interactions on mobile devices, the model was extended to include several new operators (in bold in the table below):

Table 24. From Holleis et al., 2007

Operator		Time	Quartile 1	Quartile 3
***A*, Action**	picture/marker	**1.23**	0.61	1.44
	NFC	**0.00**	-	-
	in general	*variable, input to model*		
D, Drawing		*not applicable*		
***F*, Finger Movement**		**0.23**	0.20	0.29
***G*, Gestures**		**0.80**	0.73	0.87
H, Homing		**0.95**	0.81	1.00
***I*, Initial Act**	externally	**5.32**	3.98	7.51
	internally	**3.89**	2.23	4.89
	optimal setting	**1.18**	1.10	1.26
	no assumptions	**4.61**	-	-
K, Keystroke	keypad average	**0.39**	0.37	0.48
	keypad quick	**0.33**	0.32	0.37
	hot Key	**0.16**	0.15	0.20
M, Mental Act		**1.35**	-	-
P, Pointing		**1.00**	0.84	1.20
R, System Response Time	NFC	**2.58**	2.46	2.80
	visual marker	**2.22**	2.09	2.82
	general	*variable, input to model*		
S_{Macro}, Macro Attention Shift		**0.36**	0.28	0.44
S_{Micro}, Micro Attention Shift	keypad ↔ display	**0.14**	0.14	0.19
	hotkey ↔ display	**0.12**	0.02	0.14
	keypad ↔ hotkey	**0.04**	0.02	0.12
	in general	**0.14**	0.10	0.16
***X*, Distraction**	slight	**6 %**	3 %	13 %
	strong	**21 %**	11 %	25 %

Table 1: Overview of the proposed times for all operators.

Schulz's work (2008) focused more on differences between devices and capturing KLM information with an application running in the background. The interesting findings from this work centered on applying heuristics to the M (mental thinking time) operator. For expert users of a system, M operators could potentially be placed automatically by the software, for non-expert users, it may be necessary to place the M operators manually. Schulz also introduced a new operator *I* to signify that some type of input method was being used. Schulz concluded that the KLM model was applicable to mobile phones and suggested that adding in the operators suggested by Holleis et al. would not invalidate the data generated by the model variant he employed but would provide a deeper understanding of what actually occurred.

Cox et al. (2008) used KLM to predict timings comparing the use of voice, multi-tap and predictive text entry using mobile devices. Models for each of the experimental conditions (KMK=nav: keypress + multitap, text: keypress; KPK=nav: keypress + predictive, text: keypress; KSK=nav: keypress + speech, text: keypress; SMS=nav: speech + multitap, text=speech; SPS=nav: speech + predictive, text: speech; SSS=nav: speech + speech, speech:

Tables 25 and 26. From Cox et al., 2008

KMK	$T(M) = c(T_c + T_k) + w(k_t T_k + dT_{m1})$
KPK	$T(M) = c(T_c + T_k) + w(k_p T_k + l(T_{m2} + T_k))$
KSK	$T(M) = c(T_c + T_k) + sT_s + wT_{rw}$
SMS	$T(M) = c(T_c + s_c T_s + T_{rw}) + w(k_t T_k + dT_{m1})$
SPS	$T(M) = c(T_c + s_c T_s + T_{rw}) + w(k_p T_k + l(T_{m2} + T_k))$
SSS	$T(M) = c(T_c + s_c T_s + T_{rw}) + sT_s + wT_{rw}$

Table 9 Mean time taken in seconds (and standard deviation) to complete the tasks in each condition

	Multitap	Predictive	Speech	Overall mean
Keypress	36.67 (8.53)	25.63 (7.97)	9.51 (1.85)	23.94
Speech	45.77 (11.56)	32.23 (8.48)	16.20 (2.52)	31.40
Overall mean	41.22	28.93	12.85	

Then, the KLM predicted timings were validated with empirical data. Then, a test under limited visual feedback conditions was executed to determine the effect of this on interaction. Results indicated that spoken text entry was preferred and had the fastest execution times.

In her PhD work, Luo (2008) made important strides in investigating a methodological approach to including energy efficiency as an element in user interaction design. Her work attempts to optimize efficiency by understanding use of the largest source of power consumption, the display. In so doing, Luo suggests that display use is directly related to workload, perhaps even task dependent, and understanding patterns of use can inform improved energy efficiency. Luo's work also extends the KLM model to KLEM to incorporate a prediction of system energy consumption during the task execution.

John and Suzuki (2009) conducted another model validation using a mobile phone and CogTool to predict the time needed to make a phone call from the phone book. The intention was to gather empirical data on performance, compare it with the predictive model and revise or refine the model accordingly to capture detailed information and ideally help make predictions for both skilled/expert users as well as novice users.

Expert

As noted above, the KLM is predicated on the notion of expert interaction. Because mobile devices enjoy such widespread and rapid uptake, it is assumed that expert use can be achieved rather quickly once device acquisition and use have been initiated. This assumption is especially important given that even populations with no prior computing exposure or experience are successfully using mobile devices. By reducing the cognitive load of the tasks being measured and providing adequate training time, 'expert' users can be cultivated from a pool of subjects. Though novice use of mobile devices is still interesting and provides rich information about usability, the number of non-mobile users is dwindling worldwide. Moreover, our aim is to understand, comparatively, how these systems function in terms of task execution time, all else being equal or as equivalent as possible. According to St. Amant, Horton and Ritter (2007), they defined experienced as "practiced and error-free" and their approach was to 'cultivate' expert users through practice of tasks.

7. SUMMARY

This research seeks to understand the implications of device interaction and response, in particular, display size and interaction style on information seeking behavior (ISB). In order to undertake a study looking at the question of equivalence in information searching tasks across three devices, we will investigate one important parameter: task execution time. While we'd also like to understand the cognitive components of search, in order to make comparisons across devices, our approach will be to constrain device type and searching tasks, keep cognitive load at a minimum and compare a finite measure, task execution time, head to head across devices. By focusing on a quantitative parameter, the time it takes to execute a prescribed task in a simulated work task situation (Borlund and Ingwersen, 1997), will facilitate more generalizable findings. If the who, what, when, where, and why are already determined (constrained), we'd like to understand how the 'how' varies by device. Moreover, if it can be assumed that as task complexity increases, success of information seeking decreases (Bystrom and Jarvelin, 1995), seeking modality plays an important role in success, one that should be quantifiably measurable in at least some ways.

A fundamental tenet in interaction design is the reduction of complexity (Ishii and Ullner, 1997; Chang, Gouldstone, Zigelbaum, and Ishii, 2007). Because the psychology behind each user interaction with a given computing device can be both complex and highly variable, the execution component of the process, that which is often purely device dependent/controlled, should be carefully evaluated and considered relative to existing standards for acceptable interaction. To this end, the Keystroke Level Model (KLM) of the Goals Operators Methods Selections (GOMS) techniques makes comparisons of execution times across multiple devices relatively robust and simple to observe (John and Kieras, 1996 and Bonto-Kane, 2006).

CogTool, developed as a sort of 'crash test dummy' for usability interaction prototyping, was developed by Bonnie John and others at Carnegie Mellon University (http://cogtool.hcii.cs.cmu.edu/). It can be used quickly and effectively to produce estimates of interaction times for a wide variety of interfaces and interaction styles. Our approach differs from this approach in that the aim is to reduce and constrain tasks to their most essential execution elements. This includes creating conditions in which other factors, such as connection speed, processing speed, network availability, and cognitive load can be assumed to be equivalent or minimal. It also means creating prescribed tasks which attempt to eliminate mental thinking times, system response times and other variable elements in task performance.

As the typical user becomes savvy with different mobile devices and their information search capabilities, the types of tasks being performed on a given device may shift. As the domain capacity of mobile devices increases so does the level of experience of the user increase, and, correspondingly, the level of complexity of the tasks typical users will perform using them. An understanding of the tasks being performed successfully on mobile devices and how this success compares with the traditional desktop or laptop computing environment might provide insight which could address device diversification (as in the case of the desktop/laptop environment).

In an effort to make comparisons between devices, many things must be considered, yet it is not always possible or practical to consider every aspect of a given computing device when comparing it to another. The plethora of available computing devices and their associated features is making it very difficult to make simple, direct comparisons between devices and often dilutes the potential efficacy and impact of these tools. Additionally, there may be so many considerations to account for when making comparisons that the resulting work is simply not achievable. However, because we've reached a point where mobile device differentiation is complex and user adoption is high, understanding more about how devices compare is pivotal.

Another key element in this approach is the assumption that though devices differ and display size and interaction style can vary considerably, basic functionality for certain 'primitive elements of human computer dialogue' (Jacob et al., 1993) has coalesced and user experience for those tasks can largely be assumed to be expert or cultivated to the expert level. No longer should designers be focused on the initial user of computing devices, around the globe, the dissemination of computing devices has permeated nearly every society at every level of experience/exposure. Moreover, despite differences between cultures, the fundamental tenets of interaction appear to hold true regardless of culture or context. Areas of users around the globe who have little experience or exposure to the three staid components of HCI: the display, mouse and keyboard, have much to teach us about the forms of interaction that have yet to emerge.

Current evidence surrounding multi-touch typically tends to be evaluative. New concepts for interaction are tested with a small group and evaluated often in comparison to more established methods. The advent of the iPhone yields a unique opportunity to begin to gather large scale empirical evidence about the merits of this interaction style.

Approach

In summary, our approach to data collection will be threefold: we will collect demographic information from subjects using a semi-structured interview approach. We will collect task workload data after the successful performance of each task using the NASA Task Load Index (NASA TLX) and we will collect general and overall usability information using a combination of surveys (UTAUT and MPUQ), our quantitative data collection will include the use of predictive models (Fitts's Law and the Keystroke Level Model) to predict expected task execution times. Each of these will be described in more detail below.

References

(2010). Measuring the information society. International Telecommunication Union. doi: 10.3359/oz0303157. http://www.itu.int/ITU-D/ict/publications/idi/2010/Material/MIS_2010_Summary_E.pdf

Accot, J., & Zhai, S. (1997). Beyond Fitts' Law: Models for Trajectory-Based. CHI '97.

Accot, J., & Zhai, S. (2001). Scale Effects in Steering Law Tasks, 1-8.

Aliakseyeu, D., Martens, J., & Rauterberg, M. (2006). A computer support tool for the early stages of architectural design. Interacting with Computers, 18(4), 528-555. doi: 10.1016/j.intcom.2005.11.010.

Amant, R. S., Horton, T. E., & Ritter, F. E. (2007). Model-based evaluation of expert cell phone menu interaction. ACM Trans. Comput.-Hum. In, teract141, 1-24.

Amant, R., Horton, T., & Ritter, F. (2004). Model-based evaluation of cell phone menu interaction. in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 343-350.

Arter, D., Buchanan, G., Jones, M., & Harper, R. (2007). Incidental information and mobile search. Proceedings of the 9th international conference on Human computer interaction with mobile devices and services - MobileHCI '07, 413-420. New York, New York, USA: ACM Press. doi: 10.1145/1377999.1378047.

Baeza-yates, R., Dupret, G., & Velasco, J. (2007). A Study of Mobile Search Queries in Japan. America.

Barnert, W. C. (2005). A comparison of one-handed and two-handed direct and indirect computer interaction. Techniques.

Bates, B. M. (1979). By Marcia J. Bates. Society, (July), 205-214.

Bates, M. J. (1990). Where should the person stop and the information search interface start? Information Processing & Management, 26(5), 575-591. doi: 10.1016/0306-4573(90)90103-9.

Baudisch, P., & Rosenholtz, R. (2003). Halo: a Technique for Visualizing Off-Screen Locations. New Horizons, (5), 481-488.

Bayir, M. A., Demirbas, M., & Eagle, N. (2009). Discovering spatiotemporal mobility profiles of cellphone users. 2009 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks & Workshops, 1-9. Ieee. doi: 10.1109/WOWMOM.2009.5282489.

Beard, D. V., Smith, D. K., & Denelsbeck, K. M. (1996). QGOMS: A direct-manipulation tool for simple GOMS models. In Proceedings of ACM conference on Human Factors in Computting Systems (CHI', 96.

Beard, V., Muller, E., Mauro, A., Hemminger, M., Warshauer, M., Smith, A., et al. (1993). Interpretation of CT studies: Single-screen workstation versus film alternator. Radiology, 565-569.

Bertini, E., Gabrielli, S., & Kimani, S. (2006). Appropriating and Assessing Heuristics for Mobile Computing. Methodology, 119-126.

Bila, N., Ronda, T., Mohomed, I., Truong, K. N., & Lara, E. D. (2007). PageTailor: Reusable End-User Customization for the Mobile Web. Search, 16-29.

Blackwell, A. (2010). Human computer interaction. University of Cambridge.

Blumberg, S. J., & Luke, J. V. (2009). Wireless Substitution : Early Release of Estimates From the National Health Interview Survey , July-December 2008. differences, (December 2008).

Bonto-kane, M. A. (2006). Engineering Models of Human Performance on the Design of Mobile Device Interfaces. Current, 1-34.

Brewster, S., Murray-smith, R., Crossan, A., Vasquez-alvarez, Y., & Rico, J. (2009). The GAIME project: Gestural and Auditory Interactions for Mobile Environments. Techniques.

Buranatrived, J., Vickers, P., Building, P., & Street, C. (2004). A Study of Application and Device Effects Between a WAP Phone and a Palm PDA. Purchasing, 192-203.

Burigat, S., Chittaro, L., & Gabrielli, S. (2007). Navigation techniques for small-screen devices : an evaluation on maps and web pages, (August 2007).

Burigat, S., Chittaro, L., & Ieronutti, L. (2008). Mobrex : Visualizing Users ' Mobile Browsing Behaviors. Analysis, (February), 24-32.

Butts, L., & Cockburn, A. (2001). An Evaluation of Mobile Phone Text Input Methods. Reproduction.

Buxton, B. (2008). Chapter 11: Two-Handed Input in Interaction. Analysis, 1-16.

Buxton, B. (2009). Multi-Touch Systems that I Have Known and Loved. I Can, 1-33.

Buxton, W. (2008). 1.2 Surface and Tangible Computing, and the "Small" Matter of People and Design, 24-29.

Buxton, W., & Myers, B. A. (1986). Buxton, W. & Myers, B. (1986). A study in two-handed input., Proceeding, 321-326.

Bystrom, K., & Jarvelin, K. (1995). Task complexity affects information seeking and use. Information Processing & Management, 31(2), 191-213.

Capin, T., Pulli, K., & Akenine-Moeller, T. (2008). The State of the Art in Mobile Graphics Research, (August 2008).

Card, S. K., Newell, A., and Moran, T. P. (1983). The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA.
Card, S. K., Moran, T. P., & Newell, A. (1980). The Keystroke-Level Model for User Performance Time with Interactive Systems.
Carlsson, C., Carlsson, J., Hyvönen, K., Puhakainen, J., & Walden, P. (2006). Adoption of Mobile Devices / Services – Searching for Answers with the UTAUT. Sciences-New York, 00(C), 1-10.
Chae, M., & Kim, J. (2003). An Empirical Study on the Breadth and Depth Tradeoffs in Very Small Screens: Focusing on Mobile Internet Phones. Information Systems.
Chae, M., & Kim, J. (2004). Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. Behaviour & Information Technology, (791939330). doi: 10.1080/01449290410001669923.
Chan, S. S., Lam, J., Fang, X., Brzezinski, J., Zhou, Y., Xu, S., et al. (2002). USABILITY FOR MOBILE COMMERCE ACROSS MULTIPLE FORM FACTORS. North, 3(3), 187-199.

Chang, A., Gouldstone, J., Zigelbaum, J., & Ishii, H. (2007). Simplicity in Interaction Design. Design, 15-17.

Chatting, D. J. (2008). Action and Reaction for Physical Map Interfaces. Design, 187-190.

Chittaro, L. (2006). Visualizing Information on Mobile Devices. North, 40-45.

Chittaro, L. (2008). Interacting with Visual Interfaces on Mobile Devices. Computer, 1-5.
Chittaro, L., & dal Cin, P. (2001). Evaluating Interface Design Choices on WAP Phones : Single-choice List Selection and Navigation among Cards, 1-7.

Church, K., & Cluster, A. I. (2008). Who, What, Where & When: A New Approach to Mobile Search. IUI'08, 309-312.
Church, K., Research, T., & Smyth, B. (2009). Understanding the Intent Behind Mobile Information Needs. IUI'09, 247-256.
Church, K., Smyth, B., Bradley, K., & Cotter, P. (2008). A Large Scale Study of European Mobile Search Behaviour. MobileHCI, 13-22.
Church, K., Smyth, B., Cotter, P., & Bradley, K. (2007). Mobile Information Access: A Study of Emerging Search Behavior on the Mobile Internet. ACM Transactions on theWeb, 1(1). doi: 10.1145/1232722.1232726.
Clegg, P., Bruciatelli, L., Domingos, F., & Jones, R. R. (2006). Digital geological mapping with tablet PC and PDA : A comparison. Computers & Geosciences, 32, 1682-1698. doi: 10.1016/j.cageo.2006.03.007.
Correa, B., Ishikawa, E., Ziviani, A., & Faria, M. (2008). Medical Image Analysis using Mobile Devices. Information Systems, 1380-1384.
Cox, Anna L.; Cairns, Paul A.; Walton, Alison; Lee, S. (2008). Tlk or txt? Using voice input for SMS composition, 567-588. doi: 10.1007/s00779-007-0178-8.
Curran, K., Woods, D., & Riordan, B. O. (2004). Investigating text input methods for mobile phones. Telematics and Informatics, 23, 1-21. doi: 10.1016/j.tele.2004.12.001.
Czerwinski, M., Smith, G., Regan, T., Meyers, B., Starkweather, G., Way, O. M., et al. (2003). Toward Characterizing the Productivity Benefits of Very Large Displays. Human-Computer Interaction, (c), 9-16.

Das, A., & Stuerzlinger, W. (2008). Modeling Learning Effects in Mobile Texting. Human-Computer Interaction.
Deering, M. F. (1998). The limits of human vision. 2nd International Immersive Projection Technology Workshop. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19024434.

Dervin, B. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology.

Dey, A. K., & Graphics, G. D. (1999). Towards a Better Understanding of Context and Context-Awareness.
Dunlop, M. D., & Crossan, A. (2000). Predictive Text Entry Methods for Mobile Phones. Word Journal Of The International Linguistic Association, 134-143.
Eagle, N., & (Sandy) Pentland, A. (2005). Reality mining: sensing complex social systems. Personal and Ubiquitous Computing, 10(4), 255-268. doi: 10.1007/s00779-005-0046-3.

Elezovic, S. (2008). Multi Touch User Interfaces.
Elliott, A., & Hearst, M. (2002). A comparison of the affordances of a digital desk and tablet for architectural image tasks. International Journal of Human-Computer Studies, 56(2), 173-197. doi: 10.1006/ijhc.2001.0520.
Elliott, A., & Hearst, M. A. (1820). How Large Should a Digital Desk Be ? Qualitative Results of a Comparative Study. Communications, 165-166.
Fitts, P. M. (1954). The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement, 121(3), 262-269.
Forlines, C., Wigdor, D., Shen, C., & Balakrishnan, R. (2007). Direct-touch vs. mouse input for tabletop displays. Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07, 647. New York, New York, USA: ACM Press. doi: 10.1145/1240624.1240726.

Freedman, J. B., & Henderson, J. C. (2008). Digital Usage Behavior: A Sense Making Perspective. October, 1-7.
Froehlich, J., Wobbrock, J. O., & Kane, S. K. (2007). Barrier Pointing: Using Physical Edges to Assist Target Acquisition on Mobile Device Touch Screens. Stroke.
Gray, W. D., John, E., & Atwood, E. (1992). THE PRECIS OF PROJECT ERNESTINE OR AN OVERVIEW OF A VALIDATION OF GOMS.
Han, J. Y. (2005). Low-cost multi-touch sensing through frustrated total internal reflection. Proceedings of the 18th annual ACM symposium on User interface software and technology - UIST '05, 115. New York, New York, USA: ACM Press. doi: 10.1145/1095034.1095054.
Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Human Factors, 239-250.
Hemminger, B. M. (1992). Design of Useful and Inexpensive Radiology Workstations. In S/CAR 1992, Symposia Foundation (pp. 615-624).
Hemminger, B. M., Bauers, A., & Yang, J. (2008). Comparison of Navigation Techniques for Large Digital Images. Online, 21(June), 13-38. doi: 10.1007/s10278-008-9133-0.

Hinckley, K. (2002). Input Technologies and Techniques. Text, 1-65.

Hinkley, K. (n.d.). Input Technologies and Techniques. In Human-Computer Interaction (pp. 1-52).
Hoggan, E., Brewster, S. A., Johnston, J., Interactive, G., & Group, S. (2008). Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. Methodology, 1573-1582.
Holleis, P., Otto, F., Hussmann, H., & Schmidt, A. (2007). Keystroke-level model for advanced mobile phone interaction. In Proc. CHI '07. ACM Press, New York, NY, 1505-1514.
Holleis, P., Schmidt, A., Computing, P., & Engineering, U. I. (2008). MAKEIT: Integrate User Interaction Times in the Design Process of Mobile Applications, 56-74.

Holmquist, L. E. (2006). Ubiquitous Japan. interactions, (november + december), 54-55.

Holmquist, L. E. (2006). Welcome to the mobile life! Interactions, 13(3), 57. doi: 10.1145/1125864.1125898.
Holmquist, L. E., Falk, J., & Wigstr?m, J. (1999). Supporting group collaboration with interpersonal awareness devices. Personal Technologies, 3(1-2), 13-21. doi: 10.1007/BF01305316.
Hsieh-Yee, I. (1993). Effects of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experienced Searchers. Society, 44(3), 161-174.

Huang, K. (2009). Challenges in Human-Computer Interaction Design for Mobile Devices. October, I.

Hwang, J., Jung, J., & Kim, G. J. (2006). Hand-held Virtual Reality: A Feasibility Study. Design, 356-363.
Isokoski, P., & Raisamo, R. (2000). Device independent text input: a rationale and an example. In: Proceedings of AVI 2000 Conference on Advanced Visual Interfaces, ACM, New York, pp, 76-83.
Jacob, R. J., Leggett, J. J., Myers, B. A., & Pausch, R. (1993). An Agenda for Human-Computer Interaction Research: Interaction Styles and Input/Output Devices.
James, C. L., & Reischel, K. M. (2001). Text input for mobile devices: Comparing model prediction to actual performance. In Proceedings of the ACM Conference on Human Factors in Computing Systems CHI, 365-371.
Jansen, B. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. Information Processing & Management, 36(2), 207-227. doi: 10.1016/S0306-4573(99)00056-4.
Jansen, B. J., Booth, D., & Smith, B. (2009). Using the taxonomy of cognitive learning to model online searching. Information Processing and Management, 45, 643-663. doi: 10.1016/j.ipm.2009.05.004.
Jansen, B. J., Booth, D., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. Information Processing and Management, 44(3), 1251-1266.
Jansen, B. J., Spink, A., & Pedersen, J. (2005). A temporal comparison of AltaVista Web searching. Journal of the American Society for Information Science and Technology, 56(6), 559-570. doi: 10.1002/asi.20145.

Jansen, B. J., Taksa, I., & Spink, A. (2009). Foundations of Transaction Log Analysis. Quality.
Jansen, B., & Spink, a. (2005). An analysis of Web searching by European AlltheWeb.com users. Information Processing & Management, 41(2), 361-381. doi: 10.1016/S0306-4573(03)00067-0.

John, B. E. (2004). Predictive Human Performance Modeling Made Easy. Interface.

John, B. E., & Investigator, P. (2008). CogTool User Guide. Human-Computer Interaction.
John, B. E., & Kieras, D. E. (1996). The GOMS Family of User Interface Analysis Techniques: Comparison and Contrast. ACM Transactions on Computer-Human Interaction, 4(4), 320-351.
John, B. E., & Kieras, D. E. (1996). Using GOMS for User Interface Design and Evaluation: Which Technique? Computing, 1(212).

John, B. E., & Suzuki, S. (2009). Toward Cognitive Modeling for Predicting Usability. Design, 267-276.

Jones, M. (2009). Mobile Search Tutorial.
Jones, M., Buchanan, G., & Thimbleby, H. (2002). Sorting Out Searching on Small Screen Devices. Review Literature And Arts Of The Americas, 81-94.
Jones, M., Buchanan, G., Harper, R., Xech, P., Ave, J. J., Uk, C., et al. (2007). Questions Not Answers : A Novel Mobile Search Technique. Science, 155-158.
Jones, M., G, M., N, M., K, B., & Buchanan, G. (1999). Improving Web interaction on small displays. Computer Networks, 3!11.
Jones, S., Jones, M., & Deo, S. (2004). Using keyphrases as search result surrogates on small screen devices. Small, 55-68. doi: 10.1007/s00779-004-0258-y.
Jones, S., Jones, M., Marsden, G., Patel, D., & Cockburn, a. (2005). An evaluation of integrated zooming and scrolling on small screens. International Journal of Human-Computer Studies, 63(3), 271-303. doi: 10.1016/j.ijhcs.2005.03.005.
Kamvar, M. (2004). The Role of Context in Query Input : Using contextual signals to complete queries on mobile devices. Corpus.

Kamvar, M. (2008). Query Suggestions for Mobile Search : Understanding Usage Patterns. Interface.

Kamvar, M., & Baluja, S. (2006). A Large Scale Study of Wireless Search Behavior : Google Mobile Search. Work.

Kamvar, M., & Baluja, S. (2007). Deciphering Trends in Mobile Search. Computer, 40(8), 58-62. doi: 10.1109/MC.2007.270.
Kamvar, M., Beeferman, D., & View, M. (2008). Say What ? Why users choose to speak their web queries. Work, (April 2007), 1-4.
Kamvar, M., Kellar, M., Patel, R., & Xu, Y. (2009). Computers and iPhones and Mobile Phones, oh my! Human Factors, 801-810.

Karam, M., & Schraefel, M. (2005). A taxonomy of Gestures in Human Computer Interaction, 1-45.
Karlson, A. K., & Bederson, B. B. (n.d.). Direct Versus Indirect Input Methods for One-Handed Touchscreen Mobile Computing. Human Factors.
Kelly, D. (2007). Methods for Evaluating Interactive Information Retrieval Systems with Users. Foundations and Trends® in Information Retrieval, 3(1—2), 1-224. doi: 10.1561/1500000012.
Kelly, D., & Belkin, N. J. (2004). Display time as implicit feedback: Understanding task effects. In K. Järvelin, J. Allan, & P. Bruza (Eds, 377-384.

Kieras, D. (1996). A Guide to GOMS Model Usability Evaluation using NGOMSL, (313).

Kieras, D., & Approach, T. (2001). Using the Keystroke-Level Model to Estimate Execution Times.

Kiljander, H., & Kiljander, H. (2004). Evolution and Usability of Mobile Phone Interaction Styles. Technology.
Kin, K., & Derose, T. (n.d.). Determining the Benefits of Direct-Touch , Bimanual , and Multifinger Input on a Multitouch Workstation. Text, 119-124.
Lee, S., Buxton, W., & Smith, K. (1985). A MULTI-TOUCH THREE DIMENSIONAL TOUCH-SENSITIVE TABLET. Computer, (April), 21-25.
Lin, J., Duh, H., Parker, D., Abi-Rached, H., & Furness, T. (n.d.). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. Proceedings IEEE Virtual Reality 2002, 164-171. IEEE Comput. Soc. doi: 10.1109/VR.2002.996519.
Lumsden, J., & MacLean, R. (2008). A Comparison of Pseudo-Paper and Paper Prototyping Methods for Mobile Evaluations.
Lumsden, J., Iit, N. R., & Brewster, S. (2003). A Paradigm Shift : Alternative Interaction Techniques for Use with Mobile & Wearable Devices, 1-14.

Luo, L. (2008). Designing Energy and User Efficient Interactions with Mobile Systems. Discovery, (April).
Luo, L., & E, J. B. (2005). Predicting task execution time on handheld devices using the keystroke-level model. Ext. Abstracts CHI 2005, ACM Press, 1605-1608.

MacKenzie, I. S. (1995). Movement time prediction in human-computer interfaces, 483-493.

MacKenzie, I. S. (2010). A note on ISO testing of computer pointing devices. Processing, 2-4.
MacKenzie, I. S., & Buxton, W. (1993). MacKenzie, I. S., & Buxton, W. (1993). A tool for the rapid evaluation of input devices using Fitts' law models. Systems Research, 25(1993), 58-63.
MacKenzie, I. S., & Soukoreff, R. W. (2001). Measuring Error in Text Entry Tasks: An Application of the Levenshtein String Distance Statistic, 319-320.
MacKenzie, I. S., & Soukoreff, R. W. (2002). Text Entry for Mobile Computing: Models and Methods, Theory and Practice. Human-Computer Interaction, 17(907140637), 147-198. doi: 10.1207/S15327051HCI172.
MacKenzie, I. S., & Tanaka-Ishii, K. (. (2007). Evaluation of Text Entry Techniques. Text entry systems: Mobility, accessibility, universality, 75-101.

Mackenzie, I. S. (2008). Fitts' Throughput and the Speed-Accuracy Tradeoff. Design, 1633-1636.

Mackenzie, I. S., & Read, J. C. (2007). Using Paper Mockups for Evaluating Soft Keyboard Layouts. Design.

Mackenzie, I. S., & Soukoreff, R. W. (2003). Creating a Phrase Set. Computing Systems, 754-755.

Mackenzie, I. S., & Zhang, S. (1999). The design and evaluation of a high-performance soft keyboard. In M. G. Williams & M. W. Altom (Eds.), CHI 99 conference proceedings, 25-31.

Mackenzie, I. S., & Zhang, S. X. (2001). An empirical investigation of the novice experience with soft keyboards. Behaviour & Information Technology, 20, 411-418.

Mackenzie, I. S., Sellen, A., & Buxton, W. (1991). A comparison of input devices in elemental pointing and dragging tasks. In: Proc of CHI91. ACM, New York, pp, 161-166.

Mahler, T., & Weber, M. (2008). Mobile Device Interaction in Ubiquitous Computing. Computing.

Marchionini, G. (1993). Chapter 1: Information and information seeking. Image (Rochester, N.Y.).

Marchionini, G. (2006). From finding to understanding. Communications of the ACM, 49(4), 41-46.

Marchionini, G., & Komlodi, A. (1991). Design of Interfaces for Information Seeking. Library and Information Science.

Marchionini, G., & Sibert, J. (1991). An agenda for human-computer interaction : science and engineering serving human need s. Engineering, 23(4).

Martzoukou, K. (2005). A review of Web information seeking research: considerations of method and foci of interest. Search, 10(2), 1-26.

Mclaughlin, A. C., Rogers, W. A., & Fisk, A. D. (2009). Using Direct and Indirect Input Devices: Attention Demands and Age-Related Differences. ACM Transactions on Computer-Human Interaction, 16(1), 1-15. doi: 10.1145/1502800.1502802.

Moscovich, T. (n.d.). Principles and Applications of Multi-touch Interaction, (May 2007).

Moulder, J. E., Foster, K. R., Erdreich, L. S., & McNamee, J. P. (2005). Mobile phones, mobile phone base stations and cancer: a review. International journal of radiation biology, 81(3), 189-203. doi: 10.1080/09553000500091097.

Myers, B. A. (1998). Please cite this work as : Brad A . Myers . " A Brief History of Human Computer Interaction Technology ." ACM interactions . Vol . 5 , no . 2 , March ,. Technology, 28(December 1996), 1-12.

Myers, B., Hudson, S. E., & Pausch, R. (2000). Past, present, and future of user interface software tools. ACM Transactions on Computer-Human Interaction, 7(1), 3-28. doi: 10.1145/344949.344959.

Myung, R. (2004). Keystroke-level analysis of Korean text entry methods on mobile phones. October, 60, 545-563. doi: 10.1016/j.ijhcs.2003.10.002.

Pettinati, L. (2007). Challenges of interface design for mobile devices. Retrieved from http://www.yuiblog.com/blog/2007/10/02/challenges-of-interface-design-for-mobile-devices/.

Pulli, K., Klemmer, S., Hile, H., & Broll, W. (2008). Mobile Graphics.

Qiao, L., Feng, L., & Zhou, L. (2008). Information Presentation on Mobile Devices: Techniques and Practices, 395-406.

Raento, M., Oulasvirta, a., & Eagle, N. (2009). Smartphones: An Emerging Tool for Social Scientists. Sociological Methods & Research, 37(3), 426-454. doi: 10.1177/0049124108330005.

Raghunath, M., Narayanaswami, C., & Pinhanez, C. (2003). Fostering a Symbiotic Handheld Environment. North.

Rieh, S. Y., & Xie, H. I. (2006). Analysis of multiple query reformulations on the web : The interactive information retrieval context. Management, 42, 751-768. doi: 10.1016/j.ipm.2005.05.005.

Roto, V. (2006). Web browsing on mobile phones - characteristics of user experience. Doctoral Dissertation.

Ryu, Y. S. (2005). Development of Usability Questionnaires for Electronic Mobile Products and Decision Making Methods, (July).

Saffer, D. (2009). Designing Gestural Interfaces.

Sasangohar, F., Mackenzie, I. S., & Scott, S. D. (2009). Evaluation of Mouse and Touch Input for a Tabletop Display Using Fitts' Reciprocal Tapping Task. Human Factors, 1(1), 839-843.

Schulz, T. (2007). Project Proposal—KLM Generator. Review Literature And Arts Of The Americas, (1980), 1-6.

Schulz, T. (2008). Using the Keystroke-Level Model to Evaluate Mobile Phones, (1983).

Sears, A., & Zha, Y. (2003). Data Entry for Mobile Devices Using Soft Keyboards: Understanding the Effects ; Ying Zha a of Keyboard Size and User Tasks. International Journal, (776112850). doi: 10.1207/S15327590IJHC1602.

Sharma, C. (2009). MANAGING GROWTH AND PROFITS IN THE YOTTABYTE ERA. Growth (Lakeland).

Sharp, H., Rogers, Y., and Preece, J. (2007). Interaction design: Beyond human- computer interaction, 2nd ed. New York: John Wiley & Sons.

Shneiderman, B., Byrd, D.,Croft, WB. 1997. Clarifying search: A user-interface framework for text searches. D- Lib Magazine

Shneiderman, B. (1983). Direct Manipulation: A Step Beyond Programming Languages. IEEE Computer, 16(8), 57 - 69.
Silvey, G. M., Macri, J. M., Lee, P. P., & Lobach, D. F. (2005). Direct Comparison of a Tablet Computer and a Personal Digital Assistant for Point-of-Care Documentation in Eye Care Division of Clinical Informatics , Department of Community and Family Medicine and. Family Medicine, 689-693.

Spink, A., Jansen, B., & Saracevic, T. (2002). From E-Sex to E-Commerce: Web Search Changes. English, (March), 107-109.

Spink, A., Wolfram, D., Jansen, B., & Saracevic, T. (n.d.). SEARCHING THE WEB: THE PUBLIC AND THEIR QUERIES, 1-27.

Stuart, J. (2004). TaskArchitect: Taking the Work out of Task Analysis. Solutions.
Sun, Q., & Hurst, W. (2008). Video Browsing on Handheld Devices—Interface Designs for the Next Generation of Mobile Video Players, 76-83.
Tang, F., Law, M. Y., Lee, A. C., & Chan, L. W. (2004). A Mobile Phone Integrated Health Care Delivery System of Medical Images. Methodology, 17(3), 217-225. doi: 10.1007/s10278-004-1015-5.
Tarasewich, P. (2002). Wireless Devices for Mobile Commerce: User Interface Design and Usability Peter Tarasewich, College of Computer and Information Science, Northeastern University. Idea, 1-16.
Taylor, P., & Sears, A. (2003). Data Entry for Mobile Devices Using Soft Keyboards : Understanding the Effects of Keyboard Size and User Tasks. International Journal, (917865490). doi: 10.1207/S15327590IJHC1602.
Teevan, J., Cutrell, E., Fisher, D., Drucker, S. M., Ramos, G., André, P., et al. (2009). Visual Snippets : Summarizing Web Pages for Search and Revisitation. Science.
Teo, L., & John, B. E. (2006). Comparisons of Keystroke-Level Model Predictions to Observed Data. Event (London), 1421-1426.

Teo, L., & John, B. E. (2008). CogTool-Explorer: Towards a Tool for Predicting User Interaction, 2793-2798.
Thompson, D. V., Hamilton, R. W., & Rust, R. T. (2005). Feature Fatigue: When Product Capabilities Become Too Much of a Good Thing. Journal of Marketing Research, 42(4), 431-442. doi: 10.1509/jmkr.2005.42.4.431.
Toomey, R. J., Ryan, J. T., Mcentee, M. F., Evanoff, M. G., Chakraborty, D. P., Mcnulty, J. P., et al. (2010). Diagnostic Efficacy of Handheld Devices for Emergency Radiologic Consultation. Medical Physics, (February), 469-474. doi: 10.2214/AJR.09.3418.
Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. Management Information Systems, 27(3), 425-478.
Wang, F., & Ren, X. (2009). Empirical evaluation for finger input properties in multi-touch interaction. Proceedings of the 27th international conference on Human factors in computing systems - CHI '09, 1063. New York, New York, USA: ACM Press. doi: 10.1145/1518701.1518864.

Wellner, P. (1991). The DigitalDesk b Calculator : Tangible Manipulation a Desk Top Display, 27-33.

Wellner, P. (1993). Interacting with paper on the digital desk. Communications of the ACM, 36(7), 87-96.
Wildemuth, B.M. (2009) Applications of Social Research Methods to Questions in Information and Library Science. Libraries Unlimited.
Wildemuth, B. M. (2004). The Effects of Domain Knowledge on Search Tactic Formulation. Journal of the American Society for Information Science, 55(3), 246 -258.
Wildemuth, B. M., & Morre, M. E. (1995). End-user search behaviors and their relationship to search effectiveness. Bulletin of the Medical Library Association, 83(3), 294-304.

Wilson, T. (1981). On user studies and information needs, 1-20.

Wilson, T. D. (2000). Human Information Behavior. Journal of Documentation.

Wobbrock, J. O. (n.d.). User-Defined Gestures for Surface Computing.

Wobbrock, J. O. (2006). The Future of Mobile Device Research in HCI. Access.
Wobbrock, J. O., & Myers, B. A. (2006). Analyzing the Input Stream for Character- Level Errors in Unconstrained Text Entry Evaluations. ACM Transactions on Computer-Human Interaction, 13(4), 458-489.
Wobbrock, J. O., Myers, B. A., & Aung, H. H. (2008). The performance of hand postures in front- and back-of-device interaction for mobile computing, 66, 857-875. doi: 10.1016/j.ijhcs.2008.03.004.

Wu, Y., Tao, Y., & Yang, P. (n.d.). Using UTAUT to explore the behavior of 3G mobile communication users. Framework.
Xie, I. (2009). Dimensions of tasks : influences on information-seeking and retrieving process. Journal of Documentation, 65(3), 339-366. doi: 10.1108/00220410910952384.

Xie, I., & Cool, C. (1998). The Importance of teaching "interaction" in library and information science education.
Yamashita, A. F., Barendregt, W., & Fjeld, M. (2007). Exploring Potential Usability Gaps when Switching Mobile Phones: An Empirical Study. Human Factors.
Yi, J., Maghoul, F., & Pedersen, J. (2008). Deciphering Mobile Search Patterns: A Study of Yahoo! Mobile Search Queries. Distribution, 257-266.

Yim, S., Hwang, J., Choi, S., & Kim, G. J. (2007). Image Browsing in Mobile Device Using User Motion Tracking. Human-Computer Interaction, (c), 1-2.
Yost, B., Haciahmetoglu, Y., & North, C. (2007). Beyond Visual Acuity: The Perceptual Scalability of Information Visualizations for Large Displays. Displays, 101-110.
Zhai, S., Smith, B. A., & Selker, T. (1986). Improving Browsing Performance : A study of four input devices for scrolling and pointing tasks. World Wide Web Internet And Web Information Systems, 17.
Zhang, X., Li, Y., & Blumberg, F. C. (2008). Studies on Human Computer Interface Design of Chinese Mobile Phone Users. Text, 477-485.