Mental Workload in Multi-Device Personal Information Management

Manas Tungare

Dept. of Computer Science, Virginia Tech.

Blacksburg, VA, USA.

manas@tungare.name

Manuel A. Pérez-Quiñones

Dept. of Computer Science, Virginia Tech.

Blacksburg, VA, USA.

perez@cs.vt.edu

Abstract

Knowledge workers increasingly use multiple devices such as desktop computers, laptops, cell phones, and PDAs for personal information management (PIM) tasks. The use of several of these devices together creates higher task difficulty for users than when used individually (as reported in a recent survey we conducted). Prompted by this, we are conducting an experiment to study mental workload in multi-device scenarios. While mental workload has been shown to decrease at sub-task boundaries, it has not been studied if this still holds for sub-tasks performed on different devices. We hypothesize that the level of support provided by the system for task migration affects mental workload. Mental workload measurements can enable designers to isolate critical

sub-tasks and redesign or optimize the user experience selectively. In addition, we believe that mental workload shows promise as a cross-tool, cross-task method of evaluating PIM tools, services and strategies, thus fulfilling a need expressed by several researchers in the area of personal information management. In this paper, we describe our ongoing experiment of measuring mental workload (via physiological as well as subjective measures) and its implications for users, designers and researchers in PIM.

Keywords

Personal Information Management, Mental Workload, Multiple Devices

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User interfaces – Evaluation/ methodology

Introduction & Motivation

As we amass vast quantities of personal information, managing it has become an increasingly complex endeavor. The emergence of multiple information devices and services such as desktops, laptops, cell phones, PDAs and cloud computing adds a level of complexity beyond simply the use of a single computer. In traditional single terminal computer systems, the majority of a user's attentional and cognitive resources are focused on the terminal while performing a specific task. However, in an environment where multiple devices require intermittent attention and present useful information at unexpected times, the user is subjected to different mental workload.

In an earlier study we conducted [15], users consistently reported difficulties in performing information tasks with multiple devices, especially when transitioning between/among devices. From the responses we received, we observed (from a content analysis of free-form responses) that users' adoption of various technological alternatives is guided by an innate sense of certain specific factors. We noted that several of these factors constitute mental workload, e.g. frustration level, temporal demand, and mental effort. In systems where users lacked the freedom of choice, they turned to solving problems by adopting workarounds motivated by one or more of these factors.

It has been shown that an operator's task performance is inversely correlated with high levels of mental workload [12]. Thus, we set out to explore if mental workload estimates could be used to compare task difficulty in PIM tasks. Prior work in mental workload measurement has established that physiological measures such as changes in pupillary diameter (known as Task-Evoked Pupillary Response [3]) can be used to estimate mental workload. Such continuous measures of mental workload can help locate sub-tasks of high task difficulty. Iqbal et al. [8] demonstrated that within a single task, mental workload decreases at sub-task boundaries. A fundamental goal of our research is to examine if their finding still applies when the latter sub-task is performed on a different device than the former. Our contrary hypothesis is that mental workload rises just before the moment of transition, and returns to its normal level a short duration after the transition is complete.

Systems differ in the level of support they provide for pausing a task on one device, and resuming it on another [13]. A related goal of our research is to examine if the increase in mental workload at the point of transition is correlated with the level of system support available for the sub-task of transitioning. I.e., if the system incorporates full support for task migration, we hypothesize that mental workload will be less than in case of another system where such support is lacking.

In addition, there has been no standard way to compare the effectiveness of tools, services, and techniques developed independently at different research labs. Kelly [9] notes the methodological difficulties in studying PIM because of its highly personal nature, leading to challenges in developing a set of reference tasks or cross-tool cross-task metrics. In several other task domains, workload assessments such as NASA TLX [6] have been administered instead of direct measurement of task performance metrics for several reasons: chief among them is that subjective workload assessments require less effort and instrumentation of the task, and are easier to administer. If mental workload in PIM tasks can be shown to be inversely correlated with task performance (as has already been shown in several other domains [12, 2, 5]), such a measure can be used to compare the effectiveness of these tools across varying tasks. Thus, a tertiary goal of our research is to examine whether mental workload estimates captured using the NASA TLX scale can serve as a predictor of task performance for personal information management tasks.

Related Prior Work
Mental workload is an important, practically relevant, and measurable entity [6]. The NASA Task Load Index (NASA TLX) [6] is a multi-dimensional subjective

workload assessment technique that has been applied in studies of airline cockpits [2], navigation [14], and in the medical field [5]. It combines information about specific sources of workload weighted by their relevance, thus reducing the influence of those are experimentally irrelevant, and emphasizing the contributions of others that are experimentally relevant. This reduces between-subject variability for the measure as compared to other subjective scales.

Physiological measures such as changes in pupillary diameter (known as Task-Evoked Pupillary Response) have been shown to be responsive to changes in mental workload [3] and used as a physiological measure of mental workload in several studies [7, 1]. Within a single task, mental workload decreases at sub-task boundaries [8]. Such continuous measures of mental workload can help locate sub-tasks of high task difficulty.

As the problem of information overload has worsened over the years, human attentional resources have stayed constant [11]. The issue of information fragmentation across multiple devices (the condition of having a user's data in different formats, distributed across multiple locations, manipulated by different applications, and residing in a generally disconnected manner [4]) threatens the effectiveness of users as well as of our tools and systems.

An understanding of mental workload in PIM tasks is not only expected to lead to a better understanding of why a particular tool causes high frustration or mental demand in users, but also can be used to isolate critical sub-tasks and for comparing different tools against one another.

Results from Preliminary Studies

Experimental tasks for the current study were chosen from among the most common representative tasks identified in an exploratory survey study [15] and another ethnographic investigation [16] (reported elsewhere).

File management across multiple machines stood out as the most reported problematic task. 12 out of 79 survey users said that they encountered difficulties while syncing data between multiple machines, 11 reported unexpected deletion of their data while copying across machines, and 6 reported having trouble with managing conflicting versions of files that were copied manually. Based on these findings, our first experimental task involves managing files across a desktop and a laptop, with and without support for automatic synchronization.

From the ethnographic investigation of calendar use [16], we found that paper calendars were actively used by a majority of interviewees despite the widespread prevalence of electronic calendars (corroborating the findings reported in previous studies). 35% of participants reported printing their electronic calendar for offline use. Based on this, our second experimental task is calendar management, and involves managing schedules using an online calendar and paper calendars.

From the survey, we also found that several devices are often used in groups, e.g. laptops and cell phones (reported by 52 participants), and integrated multi-function portable devices such as Palm Treos, Blackberries and Apple iPhones have begun to replace single-function devices for communication (e.g. email

and IM). Given this, we picked contact management as our third experimental task.

Methodology and Experimental Setup

This mixed-method study consists of an experiment, preceded by a questionnaire, and followed by an interview. Participants are invited to perform three tasks in two sessions each to cover three different information collections: (1) files, (2) calendars and (3) contacts. Each task is performed in two different ways in the two sessions; the difference in treatments is the level of system support for task migration. E.g. for the files task, users perform the task using either USB drives (low level of task migration support) or network drives (higher level of support.)

Each task consists of a set of instructions (between 15 and 20 each) to locate, read, modify, and save information. In each task, a few instructions include questions directly related to the information at hand. The experimenter collects the answers and uses them as a metric of task performance (details later). Interspersed within these are instructions to switch devices, e.g. one of the instructions for the file management task reads: "Complete all your work on the desktop, and prepare to travel to a different office where you will only have your laptop."

The second session is conducted (at least) two weeks after the first session, in order to minimize the learning effects caused by the first session. In this within-subjects design, ordering effects are minimized by randomizing the order of treatments between sessions, as well as the ordering of tasks within each session.

Mental Workload is measured via two different ways:

Physiological Measure: Task-Evoked Pupillary Response
Subtle yet measurable changes in pupil diameter have been associated with cognitive workload and referred to as the Task-Evoked Pupillary Response (TEPR) [3]. Participants wear a head-mounted eye-tracker throughout the duration of the experiment that permits free head movement while still tracking eye gaze and pupil diameter with reasonable accuracy. Pupil diameter (adjusted and normalized for other factors) has been shown to be a good predictor of cognitive workload [7, 10]. This technique provides a continuous measure of mental workload.

Subjective Measure: NASA Task Load Index (TLX)
After every task, participants are requested to record their subjective assessment of mental workload via the NASA TLX questionnaire. This offers a task-level estimate of mental workload that is useful as a cross-task comparison metric.

Task Performance Metrics
Direct task-related metrics such as time taken, errors encountered, information overwritten or not correctly propagated across devices, and incorrect information used are being measured and used to determine if high mental workload correlates negatively with task performance. These are measured after the participant session has concluded, by (1) analyzing eye-gaze video, (2) automatic instruction-level time-tracking in the system that displays task instructions, (3) analyzing the end products of interaction, e.g. saved files, modified calendars and (4) answers to questions posed at the end of individual instructions.

As of January 2009, pilot studies have been conducted with 8 participants and a few initial participants have been recruited and scheduled for the first session.

Expected Results & Design Implications

Designers of PIM products and services strive to create solutions that make it easier for users to get their tasks done. However, an evaluation of the effectiveness of these tools poses tricky challenges. Kelly [9] notes that "research and theory concerning PIM behavior and tools have been stymied, since it is difficult to accumulate, compare, and integrate results across studies" and expresses an urgent need for "developing evaluation methods and metrics that produce valid, generalizable, sharable knowledge about how users go about the PIM activities and interactions in their daily lives."

We believe that the results of our experiment will contribute to exactly such an endeavor. Mental workload already accounts for subjective factors such as frustration and mental demand, factors that users have reported as important in influencing their choice of device/tool/strategy. If, further, mental workload can be shown to be correlated with task performance, then it has tremendous potential in being used for cross-tool evaluations and for comparing vastly different PIM methodologies with one another. If, as we expect, we are able to find significant correlation among physiological and subjective measures of mental workload and task performance, designers will be able to evaluate their tools using non-intrusive low-overhead subjective workload assessment tests such as NASA TLX.

Not only will we be able to determine if a particular system causes higher or lower mental workload in a user, we will also be able to understand where within a task users face problems. Measures of mental workload can be used in both formative and summative evaluations of PIM products in the testing phase, and changes and/or optimizations can be introduced in case mental workload is found to be unexpectedly high during certain task sequences in a higher-level task.

Summary

In this paper, we describe a study in progress that seeks to understand the changes in mental workload during personal information management tasks performed using multiple information devices. We extend prior work in mental workload measurement to the domain of PIM, and seek to examine its correlation with task performance. Mental workload is measured via physiological as well as subjective measures, while task performance is measured using several task-specific metrics for three independent tasks (each of which was selected based on the results of two prior studies.) This study has important implications for PIM system designers who can then use mental workload measures as a cross-task, cross-tool method for comparing the effectiveness of PIM tools and services developed independently of one another.

Acknowledgments

References

[1] B. P. Bailey and S. T. Iqbal. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. ACM Trans. Comput.-Hum. Interact., 14(4):1–28, 2008.

[2] J. Ballas, C. Heitmeyer, and M. Pérez-Quiñones. Evaluating two aspects of direct manipulation in

advanced cockpits. In CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 127–134, New York, NY, USA, 1992. ACM.

[3] J. Beatty. Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin, 91(2):276–92, 1982.

[4] O. Bergman, R. Beyth-Marom, and R. Nachmias. The project fragmentation problem in personal information management. In CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, pages 271–274, New York, NY, USA, 2006. ACM Press.

[5] D. A. Bertram, D. A. Opila, J. L. Brown, S. J. Gallagher, R. W. Schifeling, I. S. Snow, and C. O. Hershey. Measuring physician mental workload: Reliability and validity assessment of a brief instrument. Medical Care, 30(2):95–104, 1992.

[6] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Human Mental Workload, 1:139–183, 1988.

[7] S. T. Iqbal, P. D. Adamczyk, X. S. Zheng, and B. P. Bailey. Towards an index of opportunity: understanding changes in mental workload during task execution. In CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 311–320, New York, NY, USA, 2005. ACM Press.

[8] S. T. Iqbal and B. P. Bailey. Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In CHI '05: CHI '05 extended abstracts on Human factors in computing systems, pages 1489–1492, New York, NY, USA, 2005. ACM.

[9] D. Kelly. Evaluating personal information management behaviors and tools. Commun. ACM, 49(1):84–86, 2006.

[10] J. Klingner, R. Kumar, and P. Hanrahan. Measuring the task-evoked pupillary response with a remote eye tracker. In Eye Tracking Research and Applications Symposium, Savannah, Georgia, 2008.

[11] D. M. Levy. To grow in wisdom: Vannevar Bush, Information Overload, and the Life of Leisure. In JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries, pages 281–286, New York, NY, USA, 2005. ACM Press.

[12] R. D. O'Donnell and F. T. Eggemeier. Workload assessment methodology, volume 2 of Handbook of perception and human performance: Vol. 2. Cognitive processes and performance, chapter Workload assessment methodology, pages 42/1–42/49. Wiley, New York, 1986.

[13] P. S. Pyla, M. Tungare, and M. Pérez-Quiñones. Multiple user interfaces: Why consistency is not everything, and seamless task migration is key. In Proceedings of the CHI 2006 Workshop on The Many Faces of Consistency in Cross-Platform Design., 2006.

[14] J. C. Schryver. Experimental validation of navigation workload metrics. Human Factors and Ergonomics Society Annual Meeting Proceedings, 38:340–344(5), 1994.

[15] M. Tungare and M. Pérez-Quiñones. It's not what you have, but how you use it: Compromises in mobile device use. Technical report, Computing Research Repository (CoRR), 2008.

[16] M. Tungare and M. Pérez-Quiñones. An exploratory study of personal calendar use. Technical report, Computing Research Repository (CoRR), 2008.