An Investigation of Memory for Daily Computing Events

Mary Czerwinski and Eric Horvitz

Microsoft Research, One Microsoft Way, Redmond, WA 98052 USA
Tel: *+01(1)425 703 4882*
Email: {*marycz, horvitz*}*@microsoft.com*
URL: *http://research.microsoft.com/users/marycz, http://research.microsoft.com/~horvitz*

In pursuit of computational tools for augmenting computer users' abilities to interleave multiple tasks, we examined computer users' ability to identify and recall computing events deemed to be important, both with and without supportive reminder tools. Memory for events occurring during computer sessions was studied both 24 hours after an initial taped session and again after a one-month period of time. Results show that memory for important computing events is fragile and that software tools could be used to augment users' memories of how they have spent their time while computing. In addition, we observed that approximately half of the events that users identified as important could be identified automatically with available computational methods, and an attempt was made to characterize the nature of the remaining events. Finally, in a probe of alternate designs for reminding systems, we found that users typically preferred to see snapshots of their computing events in a prototype reminder system, without audio, as opposed to a full video version of an event reminder system.

Keywords: Reminder systems, memory augmentation, free recall, memory prostheses, empirical findings

1 Introduction

Today's software systems and applications provide few tools for the resumption of suspended tasks. Typical pressures and distributions of tasks and communications rarely allow users to complete a single project without interruptions, and later restarts, of tasks. Portfolios of projects and goals across business and personal life require users to manage multiple suspensions, resumptions, and interleavings of efforts. The computational and graphical capabilities of personal computers hold great opportunity for assisting users with the management of multiple, interleaved tasks. However, to date, computers, and related technologies, have largely amplified the problem. Personal computers provide users with a tantalizing means for initiating multiple projects and for collecting large quantities of interrelated data and resources associated with different projects. Computers have evolved into both a general work palette, providing a one-stop shop for multiple projects, as well as a networked communication and information-tracking tool. Thus, the quantity of interleaved projects and of notifications occurring during ongoing work continues to increase. As common examples, the popularity of email and instant messaging challenges users with handling an ongoing stream of notifications, most of which may be unrelated to the task at hand. The recent blossoming of cellular phone usage, the continuing growth of wireless computing and handheld devices has ensured that users can be contacted—and disrupted—almost anywhere at any time. Recent research has demonstrated that such ubiquitous notification can be costly in that interruptions, they provide distract users from ongoing tasks (Czerwinski, Cutrell & Horvitz, 2000; Maglio & Campbell, 2000; McFarland, 1999).

Our research is motivated by our belief that personal computers provide an unprecedented opportunity to augment users' abilities to suspend and return to tasks, whether the suspension is induced by external disruptions or the result of the user-directed shifting among multiple tasks. The goal of building *memory-augmentation systems* to help users to remember and, more generally, to reinstate tasks initiated earlier, following some period of time after a disruption, can be decomposed into several subproblems. The subproblems include: (1) the automated recognition of critical events and content associated with a task, and (2) the presentation or display of components of previously suspended events or content in a manner that efficiently reminds and updates users about suspended tasks. The experiments and results described in this paper probe aspects of both subproblems.

In this paper, we review our efforts to characterize the nature of memories that users would most benefit from being reminded about. We attempted to focus our attention on the most important events to attempt to capture and to later remind users about. We framed our pursuit of these events with the question: Of the set of events in users' daily computing lives that they tend to wish to recall at a later time, which subset of events do users tend to forget? By focusing on the subset of events that are both *deemed as important* and *tend to be forgotten*, we seek to triage engineering efforts to build an augmented memory system. We hoped that such a framing of high-payoff reminders might assist us to better characterize the informational and contextual properties of the most useful reminders.

We also present experiments that probe the relative effectiveness of alternate approaches to reinstating memories of prior context via different methods for rendering previously captured events and interactions. The results on high-payoff events and on the best approach for rendering memory-jogging content provide information about two critical challenges on the path to building effective automated memory augmentations systems.

1.1 Related Work

Related research on reminder systems includes investigation of "prospective" memory (Einstein & McDaniel, 1990; Eldridge, Sellen & Bekerian, 1992; Ellis & Nimmo-Smith, 1993; Harris, 1984; Kvavilashvili, 1992; Lamming & Newman, 1992; Sellen, Louie, Harris & Wilkins, 1996; Terry, 1988; Wilkins & Baddeley, 1978). The study of prospective memory has been distinguished from traditional research on memory in cognitive psychology in that cues for retrieval are largely self-generated (one must remember to remember), and that the events to be remembered are intentions or plans to perform some action in the future. In traditional memory research, cues for retrieval typically are generated by the researcher and the memory is for events or relationships learned in the past over a very short time period in the laboratory. Prospective memory failures have been shown through diary studies to be a serious problem as reported in the daily lives of knowledge workers (Eldridge, Sellen & Bekerian, 1992; Terry, 1988), yet very little is known about the mechanisms for bringing intentions to mind, nor how technology could be used to reduce forgetting.

A recent observational study performed by Jones, Bruce & Dumais (2001) highlighted the severity of the problem of getting back to information that users had discovered previously, and the lack of existing tool support in today's software. The study explored the myriad of ways that users attempt to keep web information available for future reference. Documented strategies included emailing useful links to one's self or others for easy access via the email inbox, pasting the web information into documents, printing out web pages, saving web pages to one's hard drive, and adding the URL as a link on a personally maintained web page. They found that the "reminding" value of any one of these methods depended greatly upon a user's habits. For instance, if a user triaged their email inbox on a routine basis, the reminding value of emailing a web link to oneself is likely to be high. However, if the inbox is overloaded with email, the value of that method is likely to be low. Interestingly, browser bookmarks, a feature designed explicitly to help remind users of their important web content, were not observed to be used by the individual participants in the study, and the authors therefore concluded that bookmarks have a low reminder value. In order to boost the reminder value of bookmarks, the authors suggested integrating bookmarks into existing hierarchical organizers, such as the email folder hierarchy or the hierarchy of the file system. In addition, they suggested that bookmarks should include a better reminding function, including why the web page was relevant to the user, perhaps the context in which is was originally found, and what actions remain to be taken with the information contained on the web page.

Lamming et al. (1994) studied the usefulness of a video diary as a memory prosthesis to help knowledge workers remember work activities that were intended for the future. They demonstrated that their participants indeed forgot quite a bit about their daily activities, and that video could be useful in aiding recall of past accomplishments and future intended activities. Full video was employed as the source of reminders in this study. We find it unlikely that knowledge workers would take the time to skim through full video in order to recall past or future activities or intended behaviors. Nevertheless, the idea of capturing user activities and allowing the user to replay brief snapshots of their activities in some summarized or abstracted form could serve as the basis for a useful type of reminder system. We were inspired by this idea in the prototypes used in our study.

Renaud (2000) provided a mini-review of how interruptions wreak havoc with users' primary tasks. She pointed out that warnings can help, but that they come at a cost; many users don't necessarily go back to the primary task, or they return only after a significant delay. Renaud's discussion appears to assume that users perform tasks in a linear manner, rather than addressing the broader, common challenge of multitasking. She suggests that imagery and pictorial presentations are superior to verbal representations in aiding memory. Renaud also classifies interruptions into multiple categories. She describes visualization for context-reinstatement and reminding that attempts to support users' "limited, sequential information processing abilities." She argues that a reminder system must be an "extra"-application, easily interpretable, and linked to explanations of system states and actions. She attempts this in her system design by melding "UI sequences," selections made by the user in the user interface, to request/response sequences. Her visualization has two main components. First, there is a session history window that chunks actions into 3 groups: current, groups of 10s, and groups of 100s, with the current category as default. A textual explanation window below the clustering of actions contains a set of explanations for the system actions that are currently selected. Second, there is a "replay my actions" feature, which can perform replay actions from the beginning to the end of a system session or walk through the last *n* actions. Renaud performed a user study examining error messages and observed that users used the system to determine how to recover from errors (in the study, users had to remember system parameters entered in order to recover). Renaud focused on error recovery, rather than on the challenge of reinstating rich context after an interruption or remembering an intended activity—functionalities we desire in a reminder system.

Altmann and Gray (2000) emphasize dynamic task environments in their research, in which users have to update their task continually with new instructions. An example environment included a fighter-plane cockpit. Their results demonstrate that new instruction updating depends on forgetting old instructions, and they study how forgetting places constraints on how the new information is encoded. They label this process "preparing to forget." Altmann and Gray describe and then test key assumptions of "functional decay theory." Their data shows that, after an initial encoding phase, there is a relatively stable use phase, in which activation for an item or instruction must begin to decay (preparing to forget). In other words, after initial encoding of new instructions, a user will operate optimally if she allows those instructions to slowly decay from memory. They demonstrate, using their serial-attention paradigm, that functional decay theory can predict many aspects of how users really handle constantly updating instructions." They observe how long it takes to prepare to forget (focusing on 1 item for about 5 seconds requires an initial 1 second of encoding prior to those 5 seconds). They conclude by stating that, within dynamic task environments, if users do not have enough time to pay attention to an update and do not have enough time to let a previous task item fade from memory, situation awareness can degrade "catastrophically." Relating this finding back to prospective memory, a user cannot attend to a future behavior if the previous task is still requiring attentional resources in short-term memory. This would appear to be an important principle to keep in mind when designing technology to augment users' memory for their computing work.

Some recent work has investigated software support for users' task flow or for contextual awareness. Most of this research has either focused on visualization of calendar events or the desktop, or augmenting the user's memory via a combination of intelligent agents and wearable computers.

Lamming and Flyn (1994) developed Forget Me Not, designed as part of a wearable computing system. The system's purpose was to continuously log a user's physical context over time so that personal event retrieval could be easily accomplished at a later time. The context saved includes information about the user in both the physical and virtual worlds. Memory-aid systems like Forget Me Not tried to help with everyday memory problems such as recalling a name or finding a document. The system created autobiographies for users by capturing who they met, where they went, who they phoned, and other contextual data. The authors based their system on the concept of episodic memory, the theory that we organize part of our memory about the past into episodes (Tulving, 1977), and that the location, who was there, etc. are strong episodic cues for retrieval from long-term memory. An example of how Forget Me Not might be used includes the scenario of a user asking a note-taking application to retrieve notes from a previous meeting with the person currently co-located with the user. This prior research comes the closest in spirit to the work we are carrying out on memory augmentation, though we are extending our notions of event tracking to include thematic content as well as episodic.

The Remembrance Agent (RA) (Rhodes & Starner, 1996) also helped play the role of a memory augmentation agent by continuously displaying information relevant to the current physical context of the user. The RA differed from Forget Me Not in that it looked at and retrieved specific textual information rather than automatically summarizing the user's actions and context. RA augmented memory by autonomously presenting previous email and notes relevant to current email written by the user. Although the goal of the RA system was to remind users of potentially forgotten information, it made no attempt to discern whether or not those documents were truly likely to have been forgotten. In addition, the system required users to manually determine which documents were stored in the database for future retrieval.

Another wearable reminder system, named Memory Glasses (DeVaul, Clarkson & Pentland, 2000), uses time, location, and activity to guide its delivery of reminders to its user. It focuses on the user's context and uses sensors (a camera and a microphone) to determine the user's activity (e.g., engaged in conversation, walking, etc.). If a contact is physically near someone wearing Memory Glasses, and has a profile that matches one or more of the wearer's interests, the system can alert the wearer to that effect. Interests are limited to names, personal interests and hobbies.

Rekimoto (1999) has implemented what he refers to as "time machine computing." In this work, a user is provided with access to visualizations of what one's computer desktop looked like, at any point in time in the past, or even in the future. With this "time-traveling" desktop, it is argued that a user no longer needs a folder hierarchy. If a document is deleted from the desktop, the user can always time travel back to the date when the document was being used and retrieve it again. The basic idea is that when you travel back in time you can see the items that were open on the desktop when you were using that document of interest, thus helping you to reinstate the original past context of use for that document. One important aspect of the system is that the user can leave "reminders" to herself in the future. Although the idea is interesting, we questioned whether views of the top-level desktop alone were sufficient to reinstate context of use for many documents and tasks. We hoped to identify what it was about users' desktops that they found important in a reminder system through a study in which users' desktops were recorded and then prototypes were examined and evaluated for their usefulness. The concept of the "time machine" metaphor influenced our thinking about this problem, and in our user research we strove to understand whether or not simply capturing the state of documents and information on the desktop would provide enough contextual information to significantly enhance users' work flow and multitasking.

Despite the recent emergence of research on memory augmentation tools, users today have few tools that can assist with the remembering of tasks and content. However, researchers have found that users often develop practices with existing commercially available systems and applications to support memory needs. As an example, studies have recently shown that an email mailbox is often used as an informal to-do list (Gwizdka, 2000; Jones, Bruce & Dumais, 2001). Users will send themselves email to remind themselves of events of deliverables at a later date. One study of email tool usage showed that when reviewing their email, people often "flagged" messages that contained to do items in order to create a visual reminder (Gwizdka, 2000). In a related study we found that multiple inbox rules and strategies are leveraged to perform this reminder function, from altering the color of items to moving them to semantically labeled folders to using flags (Dantzich et al, 2002). Of course, just as with paper, these inbox reminder strategies do not proactively remind users as to their upcoming action items, so the inbox must be triaged continually. In addition, the manner in which these reminders are displayed in order to help the user remember what the reminder pertains to, its priority and context, needs to be scrutinized. Ideas for displaying such information were generated over the course of this research and will be described in a later section of the paper.

Our primary goals with this research are twofold: to determine what information computer users want to remember for future use but are likely to forget, and to construct meaningful prototypes for how to display that information to a user in a meaningful way. The area of focus for this research in the long-term is epitomized by Figure 1.

2 Experiment 1—Personal Desktop Reminding

Based on the earlier work by Sellen, Louie, Harris and Wilkins (1996), we decided to start with an in situ study of what users truly remember about desktop computing events, what they consider to be important for future awareness, and how they tend to refer to those events.

2.1 Participants

Eight participants (one female), aged 31-55, all research colleagues, agreed to participate in this experiment. Research colleagues were chosen for this initial study because of the invasiveness of the data collection; we planned to capture their daily lives on video, and to potentially include email, phone,

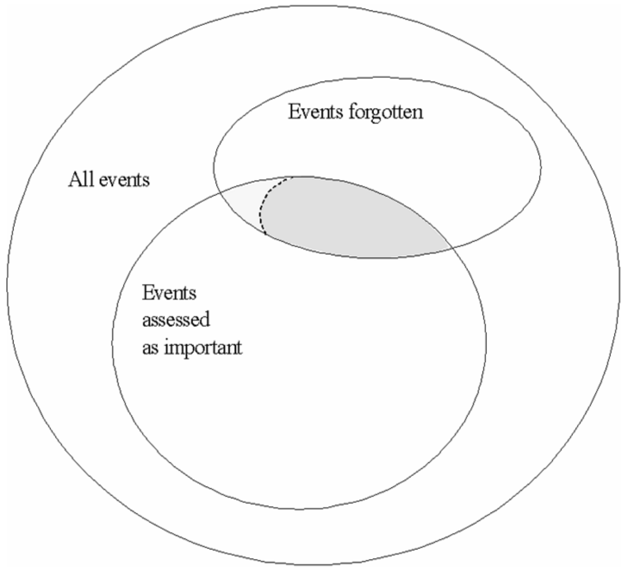

Figure 1. Characterization of event subsets. We are interested in identifying and characterizing events that are both assessed as important and that are forgotten (shaded regions at the intersection). The darker region represents the large portion of events that are important and forgotten that can be sensed with current event technologies.

hallway and instant message communications. None of the participants knew the questions motivating the study. As this study was considered a pilot study investigating memory for computing events, the sample size was purposefully small. Larger sample longitudinal studies are currently underway in our laboratory.

2.2 Methods and Procedure

Participants were scheduled for one-hour taping sessions, and were told that the video camera would probably not be close enough to capture enough detail to read their email but that regardless the content would be considered confidential. After initial camera setup, instructions to ignore the camera, and to behave as normal as possible while going about their daily research activities, approximately 50 minutes of taping was performed. For three of the participants (the first three), the experimenter was in the room with the participant, and could therefore move the camera and follow the participant around as appropriate. For the final five participants, a tripod was acquired and used in order to remove experimenter from the room so as to allow the participants to behave as naturally as possible, despite being taped. A Panasonic digital video camera equipped was used to perform the video recording.

Exactly 24 hours after the recording session, and once again exactly one month after the recording session, participants were sent an email questionnaire asking them to recall freely the events that had occurred during the taping session, for their estimates of how long these events lasted, and the order in which they occurred. In addition, they were asked to indicate which events were “higher-level” types of events, as opposed to subtasks within a goal. Users were not informed ahead of time about the existence or nature of the 24-hour and 1 month free-recall challenges. After the initial 24-hour free-recall information was received from the participants, the experimenter sat down with them individually for a half-hour video annotation session. For this session, the participants reviewed their videotapes and provided verbal commentary along with the timestamp at which the event occurred from the video camera. Subjects were told to imagine that they were reviewing their tape with an eye toward “designing a personal desktop video reminder system,” so that they could have access to *important* events that happened during that hour of taping. Participants were told to pay particular attention to key or important events that they might want to have access to in the future. The experimenter captured this commentary in her notes along with the timestamps from the video camera. This annotation session always took place within 24-48 hours of the original taping session.

Two prototypes were built from the verbally annotated videotapes. The first prototype design, which we call *PersonalStreams*, shows short video clips of the events that users’ identified as important events to be stored for future access when they had reviewed their videos. Each clip varied in length depending on what the user described as important, with a 2 second fade out and in of the next clip. The second style of prototype is referred to as *PersonalStills*, because only one representative still was captured from each video clip used in the PersonalStreams prototype. To explore the benefits or lack thereof of audio in these prototypes, we designed the PersonalStills prototypes to be silent, unless a key interaction was occurring off screen (*e.g*., a hallway conversation took place and there was no still to represent this). This was only necessary once for 2/8 participants. PersonalStreams always contained the audio that had occurred during that particular clip. On average, the PersonalStreams prototypes were 4 minutes and 44 seconds long, while the PersonalStills were on average 2 minutes and 51 seconds long (both compressed from a 50 minute session).

Because we wanted each of the prototypes to have similar duration and tempo, we needed to determine a length of a video segment that was acceptable, or a pace that felt right. In the end, we decided to base our design heuristics on the second participant's (S2's) video segmentation, as he was the most mobile. S2 often got up from his desk during the videotaping, moving from one colleague's office to another. Other times he would go to his whiteboard, or move his chair between his various computers. In order to capture the event information called out by this participant, the pace for PersonalStreams need to be more fine-grained than for the other participants. Therefore, we designed the PersonalStills prototype for S2 and then replicated that time sequencing in each of the other prototypes to maintain some consistency. Each of the stills is therefore four seconds long, including the two second display of the still, and one second each for fade in and out. We placed one second of "black" in between each of the stills. In determining the timing on the video segments, we were dependent upon the movement of the participant, letting their annotated actions set the length for how long the scene should run. Our design heuristic for the PersonalStreams prototype was to give the user enough time in each video segment of the prototype to see the key events that they annotated for us during the video review process. The one constant with the video segments was that there was always one second of "black" in between each of the segments in a PersonalStreams prototype. All prototypes were played back to participants using the Microsoft Windows Media Player, version 7.0.

After receiving email about the 30-day free recall phase of the study, participants were shown their prototypes for the first time and asked to view them twice. The first viewing was simply to watch the video. The second viewing was for participants to point out events represented in the prototypes that they had since forgotten but were reminded of by the video. Order of presentation was counterbalanced so that three participants saw the PersonalStills prototype first, and three saw the PersonalStreams prototype first. (Unfortunately, for 2 participants, the original session video quality was so low at times due to poor autofocus behavior of the camera, that only the PersonalStills prototype could meaningfully be edited). User satisfaction data was collected as well as comments and improvement suggestions about the first prototype from users before moving to the 2nd prototype.

3 Results

3.1 24 Hour Free Recall and Time Estimation

Analysis of the free recall and time estimate data within 24 hours of the initial session taping showed that participants recalled an average of 17.88 events, with a range of 5 to 40 events freely generated from memory. These data are shown in Figure 2. Participants were asked to earmark which events they thought were high level events in these free recalls, and an average of 4.63 events were called out as such (range = 2 to 8 high level events). Interestingly, on average there was one false memory (a description of an event that actually did not appear on their videotapes, given their own annotations of the tapes) per free recall summary (range = 0 to 4, mode = 0). The false memories were typically common activities that a participant usually performed at that time of day, possibly inserted because the participant had forgotten what really happened at that time. Perhaps even more surprising is the amount of overestimation observed with regard to the actual amount of time associated with events; on average participants overestimated event times by 144.9%. Finally, we found that the correlation between the order in which events were listed and their actual temporal order was quite high, r=.77.

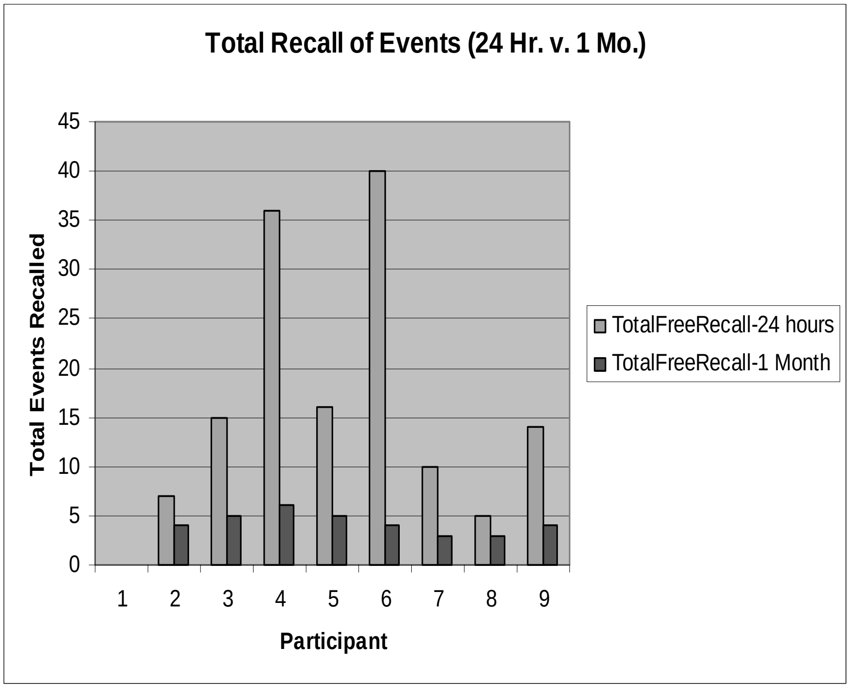

Figure 2. Total events recalled at one day and one month.

After the free-recall data was received, participants viewed their videotapes and provided verbal commentary, as described above. On average, 40.25 (range = 9 to 69) events were provided via verbal commentary by participants, providing a ratio of 0.45 of free recall/video recall events, or just under twice as many events recalled when watching the video. However, most of these events turned out to be "lower-level" events, since on average participants commented that there were only 4.75 higher-level events (i.e., events that they would truly want to have access to in the future) during the video recalls, which is a ratio of 0.98 of free recall/video recall of the higher-level events. In other words, participants had a fairly good memory for the important computing events that they would want to remember in the future after only 24 hours have gone by.

3.2 One Month Free Recall and Time Estimation

On average, participants only recalled 4.25 events from the taped session one month later, a significant drop in event free recall from the first 24 hours, $t(7)=3.08$, $p<.01$. This drop over time is shown in Figure 2. The comparison of the number of events freely recalled both after 24 hours and then one month later and the events annotated in the video 24 hours after taping is shown in Figure 3. The decrease in free-recall memory was not at the expense of new, false memories interfering with memory for events, as the average number of false memories dropped to .31 (range = 0 to 1, mode = 0), an improvement over the 24-hour free-recall performance. It also would appear that one's memory for task durations gets more accurate with some lapse of time—participants still overestimated durations one month later, but only by 26.82%, on average, though this large drop did not reach statistical significance. There was another striking decline highlighted in the one-month free recall data, memory for the temporal order of events dropped to a correlation of 0.356, a borderline significant drop when compared to the 24-hour data, $t(7)=1.96$, $p=.09$. No other differences between the 24-hour and 1-month free recall data were significant.

After the one-month free-recall data was received, participants reviewed their PersonalStreams and PersonalStills prototypes. Regardless of which reminder system was viewed first, participants remembered an average of 7.3 events after viewing the prototypes that they had not remembered in their free recall summaries, this was a borderline significant increase in event recall, $t(7)=1.89$, $p=.07$. All users benefited from the reminder prototypes, though they varied greatly in how much benefit each individual derived. One participant recalled over 50% of the desktop events one month later with the reminder system, while another participant just barely remembered more than 10% of the events even with the reminder system. Figure 4 below shows the overall benefits of the reminder prototypes for each participant one month after the initial desktop videotaping.

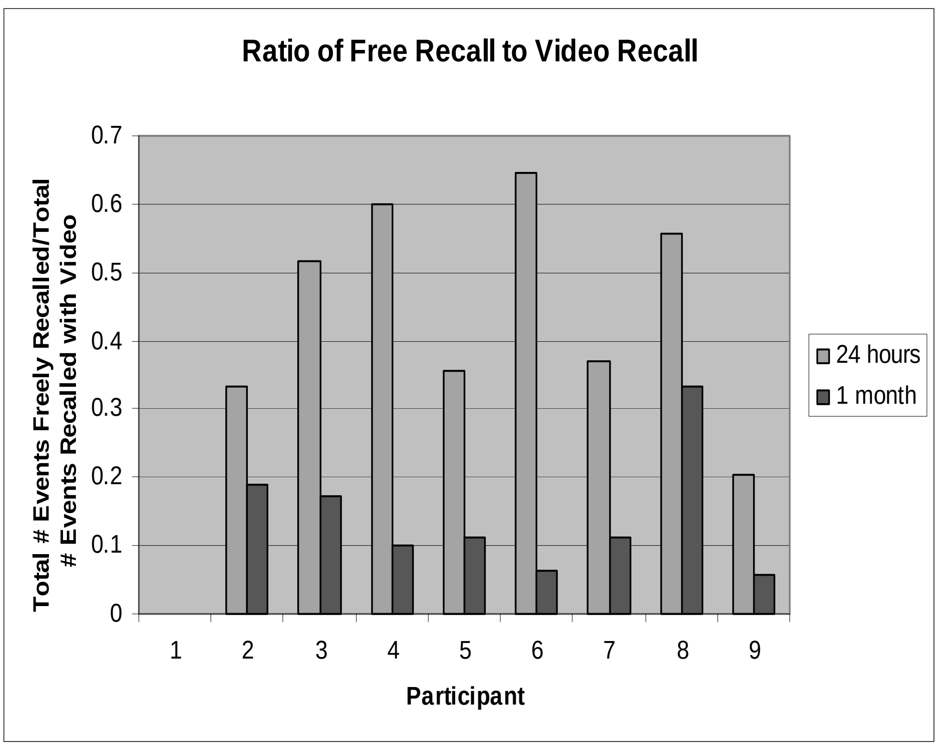

Figure 3. Ratio of free-recall to video-supported recall of events.

We did not have enough participants to examine the effects of the order of use of the PersonalStills and PersonalStreams. However, we did collect preference and satisfaction measures for each prototype that a participant viewed. Four out of six participants that received both PersonalStills and PersonalStreams prototypes preferred the PersonalStills version for its subtlety and brevity.

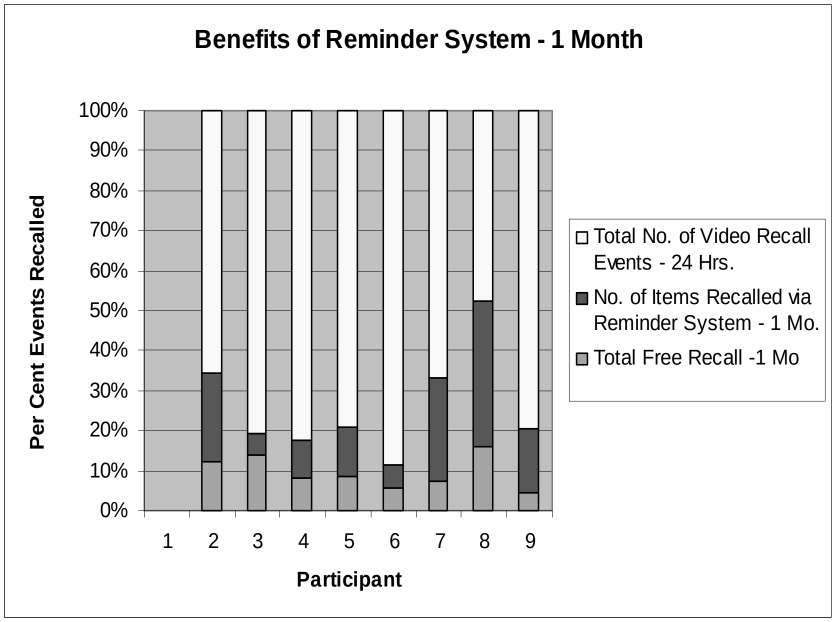

Figure 4. Boost in recall associated with use of the reminders at 1 month.

3.3 Event Identification

Event-recognition technology available today can be harnessed to identify a significant quantity of the events considered to be important by subjects. The identifiable events, which we refer to as *directly observable* events, include information about the applications that are being used or switched to and from, "patterns" of behavior, like reading and deleting email or editing a text document, and debugging and compiling code. These events and richer patterns of behavior can be detected in a straightforward manner by an event monitoring and abstraction system similar to ones developed for user modeling (Horvitz et al., 1998) or situation-assessment applications (Horvitz et al., 1999).

We found that 43% of the events identified as important by users in this study could be detected by an event monitoring system similar to that described by Horvitz et al. (1998). The system could be also be used to recognize, and store for later reminders, 22% of the 56% of the important events that had been forgotten by users within 24 hours. Events that would be challenging (though not impossible) for a system to identify include the motivations behind various patterns of behavior. Many times the user would describe the motivation behind a task switch or a behavior as the key piece of context that would help make sense of the computing event at a later time. Although user modeling applications have been making strides in reasoning about the hidden goals and needs of users, we believe that providing a means for users to efficiently annotate context in real-time or soon after interactions would be valuable in memory augmentation systems. Methods for creating annotations about context should include easy-to-use methods for capturing voice and textual comments. Such user annotation can be extended with automated event monitoring and abstraction to enhance future retrieval.

3.4 What Made People Forget Things?

Many participants reported that they forgot events that happened to them frequently, or on a daily basis. It makes sense that less important events or more mundane, frequently occurring events tend to be forgotten easily. S2 stated it best: "Things that are in my day-to-day script don't necessarily get included during the free recall exercise. I included the 'relative offsets', or information that was strongly emotional." In addition, participants described "broadcast" events as being relatively less important to keep track of mentally, *e.g.,* group email announcements were deemed less important than information that came directly to them. Finally, as reported in the memory literature, performing very similar tasks prior to the videotaped event or directly after the videotaped event interfered with the ability to freely recall events without the prototypes. Therefore, it would appear that the desktop reminder system would be especially useful to users who tend to perform the same kinds of tasks repeatedly over relatively long timeframes, for example, software developers working on large, complex systems or knowledge workers creating distinct, but similarly structured documents.

3.5 Participants' Design Ideas and Preferences

Several users shared comments and opinions about the optimal design of an intelligent reminder system. S1 said that the stills were not as "noisy," and that, qualitatively, he liked the stills much better than the full video. The stills were brief and tended to have better video quality because of the hand held camera used for taping. However, we note that S1 forgot several events and the stills were not sufficient for reminding him; he caught singular events with the stills but could not make out "back and forth" events, like an error he encountered that required him to go back and forth between two applications. S2 said that with stills he could not tell exactly what he was doing. He used the mouse movement and pointer as a clear indicator of what he was doing in the videos, especially when there were a lot of windows open. He also said he needed the audio; even hearing the key clicks was informative for him. S5 preferred stills but agreed that the audio was often more helpful than the video. S4 and S7 said they had more time to examine the events in the stills prototypes than in video, though S7 commented that the video provided more continuity. In video, however, both participants thought there was too much focus on the action, but they said that the action was not necessarily important for remembering context. S6 and S7 thought a "semantic zoom" approach would fit their requirements best. They suggested having something like a knob that first shows you still images plus some textual description of the event, but as you turn the knob you move closer and closer to veridical video and audio of the event. S6 insisted that the video would have to be very high quality in order to discern exactly what was being typed on the computer. S6 thought that seeing content being typed in documents and email was crucial for remembering context. S7 agreed, stating that many applications are purposefully designed to look alike, so being able to zoom in to the textual level of detail is very important. S6 requested the ability to stop the video at any point in time and to be able to navigate efficiently directly to that application, from the place where he was in the video (e.g., reinstating the application or system context with associated operational state). These qualitative remarks correspond fairly well with findings reported previously by Tse et al. (1998), wherein it was found that users could extract gist information from multiple streams of fairly rapid video display throughput (up to 3 frames per second for three simultaneous slide show displays), but that this high level of data throughput was "overwhelming" from a user perception point of view.

4 Discussion

We recorded a one-hour segment of desktop computing tasks from eight participants with an eye toward designing a prototype system that could remind users what they doing, or intended to do. As it was not clear at the start of this project what a user would remember and forget from that hour of recording, we collected their free recalls of the one-hour segment at both 24 hours after the initial taping and one month later. Very few psychological studies have investigated human memory for computing events, in situ, over this length of time. In addition to the free recall tasks, participants in this study verbally annotated their videotaped sessions for us so that we could make note of key events and other items of interest identified by the participants. Based on these highlighted events, we built the PersonalStreams and PersonalStills video prototypes for each individual participant, using Microsoft's Media Player as a nominal user interface for interacting with the content.

We observed that, after one month's passage of time, users forgot a significant number of events that they had deemed as important for remembering later during the original recording sessions. In addition, we found that users initially had an excellent memory for the temporal order of events, but this knowledge decayed significantly over time. We observed that the video reminder prototypes were able to augment the user's memory sufficiently about important events and about event ordering. Users remembered significantly more information after viewing the prototypes than they could generate on their own. In addition, they were reminded of the true sequence of events, which often allowed them to make better sense of their past activities.

Although participants preferred the PersonalStills prototype more than the PersonalStream prototype, some participants insisted that audio, and to some degree, video, would be required in order to remember certain events, and especially sequences of events. Based on our debriefing sessions with participants, we now believe that a combination reminder system would be ideal.

We are pursuing the augmentation of memory in several ways. On the design front, we are pursuing prototypes that provide users with a temporally linear or hierarchical list of thumbnails with textual headings of major events. We are interested in the value of audio and video playback as options for the encoding of events. Based on feedback about the prototypes studied in this paper, we believe that a user-interface metaphor similar to that provided by typical VCR or media players could suffice. In essence, participants only used the Play, Pause, Stop, Rewind and Fast Forward options to review their reminder prototypes. However, other navigation and indexing tools may be of value. As an example, one participant suggested that it would be valuable to have the ability to "jump back in time" to their document or application while viewing the video or still, and we will consider this a possible feature to add in future versions of our prototypes as well.

We also found that there is "low-hanging fruit" on the front of automation; a significant number of the events deemed as important can be detected by an event monitoring system available today. A reminder system limited to such events could be valuable in augmenting users' memories of important events. However, we suspect research will be required to limit the number of encoded events captured to the ones deemed as useful or appropriate by users, so as to minimize overloading users with too many events. We found that not all important events and context could be detected easily by current event systems. Approximately half of the events highlighted by the study participants contained a high-level intention or goal, as opposed to the application or document that they were using. These higher-level events could be supported by a reminder system if the user were provided with a means for efficiently editing or refining initial suggestions of event activity that might be provided by an automated system. Whatever the mechanism provided to support these types of events, the user interface affordance must be fairly "lightweight," so that the user could easily leave a simple reminder notation with the system prior to a task switch or in the face of an interruption.

In summary, we have reviewed several new findings in this paper on memory for events in an office setting. First and foremost, we have detailed the nature and amount of forgetting that typically occurs over the course of a day or a month about common computing events. Secondly, we have attempted to characterize the aspects of those events that an automated system would be able to identify. We also described user-interface controls that will likely be necessary in a system that can augment users' memories for typical computing events. Finally, we have provided preliminary evidence that video-based (either full or stills) reminders can significantly enhance computer users' memories for their daily tasks, especially after a full month's passage of time. We are continuing research on systems for augmenting memory, with longer-term, in situ, computing scenarios and with richer, more intelligent software reminders.

Acknowledgments

Special thanks to the participants of this study, who volunteered to be videotaped and kindly provided hours of recall data.

References

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

Czerwinski, M., Cutrell, E. & Horvitz, E. (2000a). Instant messaging: Effects of relevance and time, In S. Turner, P. Turner (Eds), *People and Computers XIV: Proceedings of HCI 2000, Vol. 2*, British Computer Society, p. 71-76.

DeVaul, R.W., Clarkson, B., Pentland, A.: The Memory Glasses: Towards a Wearable Context Aware, Situation-appropriate Reminder System. In *CHI 2000 Workshop on Situated Interaction in Ubiquitous Computing (2000).*

Einstein, G.O., & McDaniel, M.A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **16**(4), 717-726.

Eldridge, M., Sellen,A.J., & Bekerian, D. (1992). The range, frequency, and severity of memory problems at work. *EuroPARC Tech. Report No. EPC-92-129*, Cambridge.

Ellis, J.A., & Nimmo-Smith, I. (1993). Recollecting naturally-occurring intentions: A study of cognitive and affective factors. *Memory*, **1**, 107-126.

Gwizdka, J. (2000). Timely Reminders: A Case Study of Temporal Guidance in PIM and Email Tools Usage. Extended abstract in *Proceedings of CHI 2000, Human Factors in Computing Systems*, **Volume 2**, 163—164.

Harris, J. E. (1984). Remembering to do things: A forgotten topic. In J. E. Harris & P. E. Morris (Eds.) Everday Memory, Actions and Absent-Mindedness, (pp. 71-92). London: Academic Press.

Horvitz, E., Breese, J., Heckerman, D., Hovel, D., Rommelse, D. (1998). The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users, *Fourteenth Conference on Uncertainty in Artificial Intelligence* (Madison WI, July 1998). Morgan Kaufmann Publishers, 256-265.

Horvitz, E., Jacobs, A., Hovel, D. (1999). Attention-Sensitive Alerting, *Fifteenth Conference on Uncertainty in Artificial Intelligence* (Stockholm, Sweden, July 1998). Morgan Kaufmann Publishers, 305-313.

Jambon, F. (1996). Formal modeling of task interruptions. Extended abstract in *Proceedings of CHI '96, Human Factors in Computing Systems*, **Volume 2**, 45-46.

Jones, W.P., Bruce, H. & Dumais, S. T. (2001). Keeping found things found on the web. In Proceedings of *ACM's CIKM'01, Tenth International Conference on Information and Knowledge Management*, 119-126.

Kvavilashvili, L. (1992). Remembering intention: A critical review of existing experimental paradigms. *Applied Cognitive Psychology*, **6**, 507-524.

Lamming, M., & Newman, W. (1992). Acitivity-based information retrieval: Technology in support of personal memory. In F.H. Vogt (Ed*.), Information Processing '92: Proceedings of the 12th World Computer Congress*, **Vol. III**., 68-81. Amsterdam: Elsevier Science Publishers.

Lamming, M., Brown, P., Carter, K., Eldridge, M., Flynn, M., Louie, G., Robinson, P., & Sellen, AJ. (1994). The design of a human memory prosthesis. *Computer Journal*, **37**(3).

Maglio, P.P. & Campbell, C.S. (2000). Tradeoffs in displaying peripheral information. In Proceedings of *Association for Computing Machinery's Human Factors in Computing Systems, CHI 2000*, 241-248.

McDaniel, M.A. & Einstein, G.O. (1993). The importance of cue familiarity and cue distinctiveness in prospective memory. *Memory*, **1**(1), 23-41.

McFarlane, D. (1999). Coordinating the interruption of people in human-computer interaction. Human-Computer Interaction - INTERACT'99, IOS Press, Inc., The Netherlands, pp. 295-303.

Mik Lamming and Mike Flynn. Forget-me-not: Intimate computing in support of human memory. In *FRIEND21 '94 International Symposium on Next Generation Human Interface*, 1994, 1-9.

Renaud, K. (2000). Expediting rapid recovery from interruptions by providing a visualization of application activity. Proceedings of *OZCHI 2000*, 348-355.

Rekimoto, J. (1999). Time-machine computing: A Time-centric approach for the information environment. In Proceedings of the *ACM UIST'99*, 45-54.

Rekimoto, J., Ayatsuka, Y. & Hayashi, K. (1998). Augment-able Reality: situated communication through physical and digital spaces. In Proc. of the *Second international symposium on wearable computers (IEEE ISWC'98)*, 68—75.

Rhodes B, Starner T. (1996). The remembrance agent: A continuously running automated information retrieval system. In Proceedings of the First International Conference on *The Practical Application of Intelligent Agents and Multi Agent Technology (PAAM '96)*. London, UK, 487-495.

Sellen, A.J., Louie, G., Harris, J.E., Wilkins, A.J. (1997). What brings intentions to mind? An in situ study of prospective memory. Memory, 5, 483-507.

Terry, W.S. (1988). Everyday forgetting: Data from a diary study. *Psychological Reports*, **62**, 299-303.

Tse, T., Marchionini, G., Ding, W., Slaughter, L., Komlodi, A. (1998). Dynamic key frame presentation techniques for augmenting video browsing. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI) 1998*. NY: Academic Press, 185-194.

Van Dantzich, M., Robbins, D.C., Horvitz;, E. & Czerwinski, M. (2002). Scope: Providing awareness of multiple notifications at a glance. Paper submitted to *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI) 2002*. NY: Academic Press.

Weber, K. & Poon, A. (1994). Marquee: A tool for real-time video logging. In Proceedings of *Human Factors and Computer Systems CHI '94*, 58-64.

Wilkins, A.J. & Baddeley, A.D. (1978). Remembering to recall in everyday life: An approach to absent-mindedness. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), *Practical aspects of memory*. London: Academic Press.