Personal long-term memory aids

by

Sunil Vemuri

B.S., Cybernetics, University of California, Los Angeles, 1992
M.S., Computer Science, Stanford University, 1994

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology
[February 2005]
September 2004

Author:____________________________
Program in Media Arts and Sciences
September 20, 2004

Certified by:____________________________
Walter Bender
Senior Research Scientist
Program in Media Arts and Sciences
Thesis Supervisor

Accepted by:____________________________
Andrew Lippman
Chair, Departmental Committee on Graduate Students
Program in Media Arts and Sciences

Doctoral Dissertation Committee

Walter Bender
Senior Research Scientist
MIT Program in Media Arts and Sciences

Christopher Schmandt
Principal Research Scientist
MIT Media Laboratory

Rosalind W. Picard
Associate Professor of Media Arts and Sciences
MIT Program in Media Arts and Sciences

PERSONAL LONG-TERM MEMORY AIDS

by

SUNIL VEMURI

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning on September 20, 2004 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Media Arts and Sciences

Abstract

The prevalence and affordability of personal and environmental recording apparatuses are leading to increased documentation of our daily lives. This trend is bound to continue and it follows that academic, industry, and government groups are showing an increased interest in such endeavors for various purposes. In the present case, I assert that such documentation can be used to help remedy common memory problems.

Assuming a long-term personal archive exists, when confronted with a memory problem, one faces a new challenge, that of finding relevant memory triggers. This dissertation examines the use of information-retrieval technologies on long-term archives of personal experiences towards remedying certain types of long-term forgetting.

The approach focuses on capturing audio for the content. Research on Spoken Document Retrieval examines the pitfalls of information-retrieval techniques on error-prone speech-recognizer-generated transcripts and these challenges carry over to the present task. However, "memory retrieval" can benefit from the person's familiarity of the recorded data and the context in which it was recorded to help guide their effort.

To study this, I constructed memory-retrieval tools designed to leverage a person's familiarity of their past to optimize their search task. To evaluate the utility of these towards solving long-term memory problems, I (1) recorded public events and evaluated witnesses' memory-retrieval approaches using these tools; and (2) conducted a longer-term memory-retrieval study based on recordings of several years of my personal and research-related conversations.

Subjects succeeded with memory-retrieval tasks in both studies, typically finding answers within minutes. This is far less time than the alternate of re-listening to hours of recordings. Subjects' memories of the past events, in particular their ability to narrow the window of time in which past events occurred, improved their ability to find answers.

In addition to results from the memory-retrieval studies, I present a technique called "speed listening." By using a transcript (even one with many errors), it allows people to reduce listening time while maintaining comprehension. Finally, I report on my experiences recording events in my life over 2.5 years.

Thesis Supervisor: Walter Bender

Title: Senior Research Scientist

Acknowledgements

There are many reasons a student can offer thanks to an advisor: the opportunity to learn, sound guidance, funding, and, of course, offering admission. Walter Bender did all of this and more; I could not have asked for better advisor. Though Walter and I never seemed to agree on some terribly important issues (Yankees vs. Red Sox, MacOS vs. Windows, living in Boston vs. California), this did not get in the way of us working on memory aids. Walter enjoys the interesting problems, is not afraid if they are hard, and always seems to maintain his easy-going approach.

I also thank the other members of my committee: Chris Schmandt and Rosalind Picard for their insightful and timely suggestions and keen perspective on the tough issues.

Thanks to all members of the Electronic Publishing group, past and present. There are many names to mention here, and I want to highlight a few that helped in immeasurable ways: Erik Blankinship, Dan Gruhl, Jon Orwant, Vadim Gerasimov, and Joan Morris DiMicco.

I also wish to thank the Speech Interfaces group who, aside from all the research-related assistance, made me feel as welcome in their group as I do in mine.

Special thanks to Brad Rhodes, Natalia Marmasse, Jacky Mallet, Henry Lieberman, Ted Selker, and Pattie Maes.

Thanks to Felice Gardner, Linda Peterson, Pat Solakoff, and Meg Kelly-Savic for making all the big bumps in the road smaller.

Many talented UROPs stood in the trenches with me to help see this work to fruition: Philip DeCamp, Stefanie Tellex, Brad Lassey, Alisa Marshall, Sherry Zhao, Paul Soto, Tushara Karunaratna, and Dominick Rabiej.

There are many other members of the Media Lab community (past and present) who have provided guidance, assistance, and friendship over the years: Jack Driscoll, Henry Holtzman, Mike Bove, Matt Tragert, Paula Aguilera, Jim McBride, Rick Borovoy, Brian Silverman, Fred Martin, Stefan Marti, Mitchel Resnick, Alexandra Kahn, John Archambault, the Necsys group, and the Media Lab Student Committee.

I've had the pleasure of discussing this work with some distinguished researchers outside of the Media Lab who have provided valuable input into this work: Mik Lamming, Anthony Wagner, Elizabeth Loftus, and Charles Nesson.

Without the lab's sponsors, there would be no Media Lab. The funding is important, but the ability for students to interact directly with sponsors and build relationships was equally, if not more, significant to this work. In particular, I wanted to highlight a few representatives from sponsoring companies who influenced and helped this research: Dave Boor, Tom Gardos, Chalapathy Neti, and Edward Epstein.

Several notable people were influential in my career and my thinking prior to coming to M.I.T.: Rao Machiraju, Mike Graves, Richard Fikes, Bob Engelmore, Tom Gruber, James Rice, Adam Farquhar, and Terry Lee.

Thanks to all the experimental subjects.

Thanks to Deborah, Carol, and Pat for always greeting those who work late nights in the lab with a friendly face and a smile.

Of course, thanks to my wife Lakshmi, my parents Venkateswararao (Rao) and Uma, and my sisters Sita and Mytilee.

Finally, I offer thanks and apologies to all of those whom I forgot to mention here. Unfortunately, my memory aid was not very helpful when I wrote this section.

Table of Contents

List of tables

List of figures

Chapter 1 Introduction

> *I get mad at myself when I'm sitting there trying to write, and I want to recall a specific statement, a specific fact, a name, and it doesn't come immediately. I hate to research something that ought to be right there at the press of a little button in my mental computer.*
>
> - Walter Cronkite [90]

Imagine being able to reminisce about one's childhood with additional vividness and clarity. Imagine students listening diligently to a lecturer instead of feverish note-taking while maintaining (or even improving) their ability to absorb the material. Imagine being able to remember people's names better. Imagine being able to remember a perfect quote, anecdote, or joke at just the right moment in a conversation.

This dissertation is about human memory and how computers can help people remedy common, everyday memory problems. Human memory is generally poor, prone to error and to manipulation. It fails in unpredictable ways, at the most inconvenient times, and sometimes, with dire consequences. Despite this, most people function by conceding to memory fallibility.

Faulty memory can have societal impacts: for example, memory failures often result in time and resource wastage. Memory errors, both individual and organizational, are often blamed for airline, environmental, and industry disasters [55]. Innocent lives have been ruined as a result of faulty memory conveyed as eyewitness testimony in courts of law [70]."People are poor at remembering meetings and even less at remembering rationales." [97] In order to learn from your mistakes it is necessary to remember them.

Memory aids have existed for as long as one can remember: strings tied on fingers, Post-it Notes™, and Palm Pilots™ are all examples. These aids can be characterized by either being mechanisms for reminding or being off-body repositories for the explicit storage of memories for subsequent retrieval.

A new type of memory aid is now possible, facilitated in large part by low-cost miniaturized computing devices that can store a lifetime's worth of data coupled with retrieval techniques to make such archives a useful repository for memory assistance. It is this new generation of memory aids that is the subject of this dissertation.

The idea of recording everything in one's life is not new. An early proposal, Memex, dates back to 1945 [18]. More than a decade ago, ubiquitous audio-recording systems were built and studied using desktop computers and workstations [52]. Now, improved portability facilitates increased presence and it follows that more and more industry, academic, and government groups are investigating the possibility of recording everything in one's life [42,43,64,65,66].

But, recording everything is the easy part. The outstanding challenge is turning vast repositories of personal recordings into a useful resource while respecting the social, legal, and ethical ramifications of ubiquitous recording. This dissertation examines amassing data for the purpose of helping people remedy common, everyday memory problems

1.1 A true story

Over a year ago, a fellow student and I spent an evening hardware-hacking an iPaq (a handheld computer) so it could accept an external microphone. This was an esoteric, procedure that involved a good deal of trial-and-error; there were no instructions in a book, on the web, etc. We knew someday we might want to refer to what we had done, so we took notes, we took pictures, and we even wrote up a document explaining what we had done and published it on the web.

Recently, that same student asked me about something related to that hack. It had been a long time and we both forgot. So, we looked at the document we wrote, but the answer was not there. It was neither in the notes, nor in the photos. Way back, we were commending ourselves for meticulously documenting our work since we anticipated that it would be valuable later. But, we missed something and we wanted to remember now.

If it were important, we could try to find some other way of figuring it out again. If it were really important, we might even open up the iPaq again. Otherwise, we would have to give up because there was really nowhere else to look.

But, back when we were hacking, we did one other thing. We audio recorded the entire session; there was one more place we could look. It was several hours of audio, and we did not feel like listening to all of that to find just a short bit.

Fortunately, the computer software I wrote as part of this research project allows me to quickly search through all of that audio. With this software, we were able to find the answer within a few minutes.

1.2 Approach

An ideal memory aid would proactively determine when memory problems occur and provide just-in-time, minimally intrusive remedies. One way to accomplish this might be through an artificial neural interface (similar to a technology several generations beyond today's state of the art, the BION [100]) that continuously monitors a person's thought processes and interjects timely memory assistance. A less-invasive approach might include external sensors that monitor a person's actions, environment, speech, and emotional state; when a memory problem is detected, the most-recently monitored activity could be used to direct "situated" information queries whose results are presented via an eyeglass-worn computer display. Such visions shall safely remain, for the moment, the domain of science fiction.

1.2.1 Solving memory problems

Conventional remedies to forgetting depend upon the type of memory problem encountered. For memory problems in which the person is aware of the problem when it occurs, an active approach might be taken. For example:

- Ask someone else (in person, telephone call, instant message, email, etc.) [a "lifeline"]

- Look it up (book, encyclopedia, web search, email search, etc.)
- Retrace steps (if you lose something)
- Continue despite the problem (common for tip-of-the-tongue forgetting)

Other types of memory problems occur without the person's awareness. For example, forgetting to buy eggs while at the grocery store, or forgetting to pay a bill. When one *anticipates* such lapses, the typical strategy is prevention. For example:

- "To do" lists
- Strings on fingers, writing on one's hand, or other bodily reminders
- Setting alarms
- Requesting reminders from other people

Computers can assist with all of these strategies, and to a certain extent, already do. For example, when trying to remedy a problem by asking someone, people use computers as communications channels (email, instant messaging). Software packages offer reminder and alarm mechanisms based on calendars, "to do" lists, etc. Computers are also effective tools for looking up information (e.g., on-line encyclopedias, web search tools, email search tools, etc.).

The memory problems and strategies of present interest are only those in the first list above and some of the limitations of these are described below. Co-witnesses may forget details of past events. Even if they claim to remember, they may lack confidence in their own memory or falsely believe they are remembering correctly. This can be particularly harmful in legal eyewitness-testimony situations as studies reveal unjustifiably high value placed on such evidence [70].

The benefits of "contextual dependency" (Section 2.1.6) explain why retracing one's steps can serve as a powerful memory trigger. But, this strategy requires remembering one's steps and the memory problem must be something that can be remedied via context.

Resources that one can "look up" are only useful if the data are documented. For example, historical facts, well-known scientific principles, literary works, written opinions, and similar information are documented in some form and can be referenced. Public and semi-public events are often documented and published by stenographers, journalists (professional and amateur), tourists, and other voyeurs. But, even the most vigilantly recorded events have gaps. In a courtroom setting, clients and attorneys may whisper privately to each other; journalists and stenographers may miss potentially interesting moments. For example, in June 2004, Vice President Dick Cheney uttered a profanity-laden phrase to Senator Patrick Leahy. This attracted considerable press attention, but the exchange occurred during an U.S. Senate photography session—outside of the normal senate proceedings. No stenographer was capturing it and no reporter had recorded the utterance; the actual phrase was left to speculation. Even if this happened in a congressional proceeding, the Congressional Record can be amended after the fact to eliminate gaffs. In some instances, some may think an event is recorded and will be available later, but for various reasons, is not. For example, the infamous "18-minute gap" found among President Richard Nixon's oval office audio recordings.

Though we may intersect with a documented situation from time to time (e.g., security cameras, public events), our daily lives—for the most part—are less-documented than the examples above. Nevertheless, they are filled with communications that might be useful for future memory problems. Reder and Schwab [85] studied workplace communications in Fortune 500 companies and found that, depending upon one's role (senior management, marketing, and sales personnel), 48–84% of the day is spent in some form of communication (face-to-face, phone, and "other") with the remainder dedicated to solitary work (the study was done prior to pervasive e-mail).

Research interest in meeting room [77,78] and lecture hall [1] archival suggests some workplace and classroom communications will be documented more frequently. Memory problems related to events that transpire in these situations can be serviced using these records. But, this still leaves a multitude of undocumented and possibly more-private events: conversations in offices, in hallways, in recreation areas, in eating venues, in shared lounges, near beverage dispensers (e.g., the coffee machine or water cooler), etc. Also, these efforts do not address personal conversations with one's friends and family in various settings (e.g., homes, parks, pubs, etc.).

A verbatim, unimpeachable record of every experience can resolve much of these issues and, again, computers can help.

1.2.2 The iRemember "memory prosthesis"

The technology introduced herein attempts to enhance a person's ability to remedy memory problems by *computer-assisted active look-up* of past experiences when the person recognizes the problem is happening. The approach is two-fold. First, one must have data to look up; a computer attempting to help trigger memories is well-served by having a comprehensive digital record of past experiences. Since my interest is in long-term memories in real-world situations, I have created a wearable recording apparatus that enables recording in such situations. Details of this device are presented in Section 3.3.1. With this, I have accrued 2.5 years of selected experiences (including verbatim audio recordings) and associated contextual data. This is one of the most-comprehensive long-term individual-experience collections accrued to-date.

At present, creating such an archive is an uncommon venture requiring vigilance beyond what most would be willing to offer. Suffice to say that investigators with research agendas are dominant among the "diligent archivers" discussed in Section 2.4. One thing I share with this group is the belief in the inevitability of ubiquitous recording and the utility of personal data archival for *something*. This dissertation examines the value of such archival towards remedying memory problems.

Collecting data is not enough. One must also have the means to find memory triggers within this archive to remedy the problem at hand. "Computational memory retrieval" will be discussed in Section 1.3 and the present tools designed to assist with "memory-retrieval" are described in Chapter 3. Some of these techniques are already used in other areas such as information retrieval, but have been refined to the present memory-retrieval task. Chapter 4 includes details on the evaluation of these techniques towards remedying selected memory problems.

1.2.3 Data capture

A design goal is to minimize the effort needed by people to capture many daily experiences. This means, when possible, the data capture tool should be available in all desired situations and data should be captured with little effort or passively. To this end, a wearable solution is advocated.

Capturing as much as possible about past events is a somewhat brute-force approach that minimizes the chance of missing potentially valuable memory triggers at the cost of retaining superfluous ones. This approach is enabled in large part by low-cost high-capacity digital storage. In effect, it transforms the memory-triggering problem into one that is not particularly cunning, but well-suited to computers: search. Interest in continuous archival of personal experiences is growing (Section 2.4). As more become involved in the collection of such data, other techniques to make use of these are likely follow.

An alternate approach would be to distill events down to a limited set of individually tuned triggers. Even if the necessary relevance assessments and deductions for such an approach could be accomplished, any assessments at the time of capture may not reflect all possible future needs. For now, it is simpler to retain everything and assess relevance on a per-need basis. The brute-force approach is chosen as a reasonable first-pass at the problem.

It should be noted that completely passive data capture may in fact hurt memory recollection as evidenced by the disadvantage of no note-taking among students in classroom situations [76]. The choice to prefer passive data capture is to reduce forgetfulness in the daily data-capture process. Ironically, forgetting to activate the memory prosthesis (i.e., absent-mindedness) is a common problem.

1.2.4 Data sources

Many things can act as memory triggers: a photograph, the smell or taste of food, someone's voice, the cheer of a crowd, etc. Different people experiencing the same event will likely retain different memories and would need different triggers to elucidate the forgotten or clarify the misremembered parts.

Since I have advocated the record-as-much-as-possible-and-hope-it-helps-later approach, the next question is what data sources can and should be captured. An ideal data source has maximal memory-triggering value while presenting minimal computational and storage demands.

Audio

High on the list of desired data sources is audio, due to the anticipated memory-triggering value of speech. Speech is rich in content and includes emphatic cues like prosody and cadence that add meaning and subtlety. Speech is pervasive and can be captured easily in a variety of settings using low-cost portable devices. Most commercially available portable and handheld computers today come with built-in audio-recording capabilities; more and more public and semi-public environments (e.g., meeting rooms, auditoriums, classrooms etc.) are outfitted with audio-recording facilities. Mobile phones (with an estimated 1.35 billion in use today [45]) and some portable music players now offer audio-recording options.

A variety of speech processing tools, such as speech recognizers, that can transform speech into a more-easily searchable representation are readily available. The data storage requirements are tractable; for example, one year of audio (assuming roughly 20 hours per week) compressed at 10:1 could be stored in roughly 20 gigabytes on a commercially available computer hard drive for around US $30.

However, both ubiquitous audio recording and archival have privacy, social, and legal implications. Depending upon the local laws [51] and social conventions audio recording may require consent from all participants for each recording and collecting audio may neither be completely passive nor continuous.

Other sources

Similar to doppelgänger [83], sources that are captured and archived both continuously and passively include the user's location, calendar, email, commonly visited web sites, and weather. Still photography is a common request; capturing biometrics [43] was also considered and both are items for future work.

Video was considered. In the early design stages of the project, the hardware complexity of a wearable, continuous-video-capture device combined with the data storage requirements and difficulty of analyzing and indexing hundreds of hours of video suggested otherwise. The field has changed rapidly since then. Research on wearable video-capture [127] and retrieval systems [102] has gained momentum with some promising results. This may soon extend beyond research environments as mobile phones with photo- and video-capture capabilities are become more commonplace. Storage capacities in portable devices have increased dramatically and improved video compression techniques are reducing the storage demands. It would not be surprising to see a similar video-based long-term memory aid in the near future.

1.2.5 Applications

This research is aimed at assisting healthy individuals with normal, everyday memory problems. Within this population, the research rests upon the intuitive notion that memory problems are generally bad, remedies to these would be good, and some everyday tasks could be performed better with fewer memory problems. One of the goals of this research is to identify the tasks subjects choose to use iRemember of their own volition.

This research does not prescribe these tasks *a priori*. Instead, the approach is meant to reveal the applications and tasks that are advantaged by the memory prosthesis through everyday use *in situ*. It is hard to predict what tasks subjects would select since such vigilant, detailed recording is an uncommon practice today. Only a limited set of people—mostly researchers—has shown willingness to engage in such behavior, accrued personal data archives, and can provide testimonials towards the benefits. By building experimental prototypes that can be given to non-investigators (albeit sympathetic colleagues), I hope to broaden the applicability of the results beyond investigators. Experiences of investigators and non-investigators who have used the memory prosthesis for daily recording are reported in Section 4.5.

Outside of individual memory assistance, there are certain tasks and occupations that are intrinsically well-suited to a ubiquitous-recording continuous-archival approach by virtue of the vigilant recording that already takes place. For example, news reporters regularly tape interviews and review recordings when writing articles. Ethnographers often record subject interviews and also review for data analysis and scientific article writing.

1.3 Computational memory retrieval

The well-studied field of *information retrieval* investigates how to build computer-based systems to find information relevant to a user need within large collections. The SIGIR conference, held annually since 1977, solicits publications on techniques to improve information-retrieval systems. The Text Retrieval Conference (TREC), held annually since 1992, brings together systems designers to compare system performance via a set of objective criteria (e.g., precision, recall, mean-average precision) using standardized corpora, queries, and tasks.

The historical focus for both conferences has been on retrieval of English text documents based on text queries. Over the years, many "tracks" formed studying related or more-focused sub-problems. The TREC spoken document retrieval track (SDR) discussed in Section 2.3 is one example. Others include multimedia information retrieval, foreign language document retrieval, and question-answering tracks. These are all well-studied with established communities, refined experimental methods, sample corpora, objective metrics, etc.

Given the thesis' emphasis on memory, I now define memory retrieval.

> *Memory retrieval leverages information-retrieval techniques for the purpose of finding information that, in turn, solves memory problems.*

With this introduction come some questions and preliminary conjectures:

- What data would one search to remedy memory problems?

  In theory, any data that remedies a memory problem is sufficient. But, the data collected from personal experiences are anticipated to have the most value. This can include anything we say, hear, see, read, etc. throughout the course of our lives.

- What are the success metrics for memory retrieval?

  Unlike other information retrieval tasks, the criteria for memory-retrieval success is not based on finding a specific piece or pieces of information, rather it is based on finding a memory trigger within a collection that remedies the desired memory problem. This is probably person- and situation-specific. One piece of information may work for one person while not for another. For the same person, a particular piece of information might trigger a memory in one instance, but not in another.

Hence, unlike precision, recall, and mean average precision, the success metric for memory retrieval is based on human performance instead of a mathematically grounded information-theory-based computation.

- How effectively can memory problems be addressed via information-retrieval techniques applied to a personal-data archive?

  Unknown. Differences are expected among the search *strategies* employed by those trying to remember a previously witnessed event versus someone trying to find information within an unfamiliar collection; in the former case, retrieving any information that triggers the memory of the event is sufficient; in the later, finding the exact information is necessary. When the data or task resembles an existing information-retrieval task, one would expect similar performance. For example, if the personal data included speech, one would expect some of the lessons of SDR to apply.

- What customizations can be made to information-retrieval techniques to improve memory-retrieval performance?

  Unknown. In fact, this is something that this dissertation hopes to explore. In SDR, empirical studies on various customized algorithms to have yet to demonstrate ranking benefits [99]. Despite these findings and the strong reliance on SDR for the present task, customizations are expected to benefit memory retrieval. Since the data are intertwined with a person's biological memory, ranking that leverages episodic context, biometric salience, or person-specific-relevance may help.

- To what extent does a person's memory help or hurt memory retrieval?

  Unknown. However, if the process of memory retrieval can be supplemented by the person's accurate recollection of related information and circumstances, their memory is expected to be beneficial towards memory-retrieval. If their memories are inaccurate, it may hurt performance.

- Are there quantitative and qualitative differences between memory-retrieval performance and strategies across different memory problems?

  Unknown. Different memory problems are expected to need different solutions. With regard to the memory problems of interest, it is not clear if some are more-benefited by the memory-retrieval approach than others.

1.4 Research questions

The thesis of this dissertation is:

> *Long-term personal-data archival can be effectively used for memory assistance.*

Intuitively, having a detailed record of a past event would be a valuable and unimpeachable resource towards resolving memory problems. However, when mixed with hundreds if not thousands of other recordings, memory remedy becomes an overbearing problem of finding the proverbial "needle in the haystack."

With enough time and incentive, people could probably remedy most memory problems via a comprehensive archive of past experiences. Since most real-world situations do not provide either such time or incentive, the primary goal of the memory aid is to remedy problems within the constraints of attention, time, and effort people can and are willing to commit. The computer-based information-retrieval approach is known to work with adequate speed and accuracy for queries on large text corpora. It is hypothesized that a similar approach will work adequately for memory-retrieval tasks using the aforementioned data sources (Section 1.1.4).

Under the umbrella of the thesis above are more-specific research questions:

1. What technologies prove most-useful?

> One can propose a multitude of approaches to co-opt personal-data archives for memory assistance. The present information-retrieval approach is but one. Even within this scope, different technologies along the chain (e.g., data analysis, information extraction, indexing, user interface, etc.) can affect the overall value of the recorded data towards memory repair. In the present work, some technologies have been borrowed and tweaked, others have been newly invented, but all have been offered to subjects to see what works best for the present task.

2. What memory problems are well-served by this approach? Which ones are not?

> A taxonomy of memory problems and their frequencies is presented in Sections 2.1.3 and 2.1.4. The active-lookup approach necessarily eliminates some memory problems from consideration. For example, problems in which the sufferer is typically not aware of the problem when it occurs (e.g., absent-mindedness) and would not pursue a remedy at the time of need. Other "forgetting" memory problems in which the sufferer is aware of the problem and wishes to pursue a remedy are better fits to the present approach. The evaluations discussed in Chapter 4 quantitatively and qualitatively examine the value of the memory-retrieval aids for these problems.

2. What data sources prove most-useful?

> At the risk of spoiling the punch line, it is probably no surprise to the reader that the memory aids described are able to assist with memory problems. This and the next research question add color to this.
>
> Existing memory research has already provided some insight into this question for audio via laboratory memory studies [126]. Although cues like smell are

particularly resistant to forgetting [34], it is not included as is part of this research for reasons stated earlier (Section 1.1.4). Among the data sources that are recorded, are some more valuable than others? Are there differences between what is found in the laboratory memory studies and what is found when examining personal-experience data?

4. What specific tasks would people use such memory aids?

> As mentioned in Section 1.1.5, there is not a specific memory-demanding task driving this research. Instead, the investigation aims to uncover the tasks that people naturally choose when given a memory prosthesis or other similar ubiquitous-recording devices. Anecdotal data based on real-world use are reported in Section 4.1.

5. What are the social implications? What social norms does ubiquitous recording challenge?

> The ubiquitous-recording premise includes a not-so-subtle implication of scores of people carrying devices and maintaining permanent archives of day-to-day experiences. If even partially realized, the situations recorded would go far beyond the sporadic images and audio captured by tourists, photographers, videographers, and other voyeurs today. There are inescapable social and privacy implications. Section 5.2 discusses these in more detail

6. What are the legal implications?

> Having a personal archive of daily experiences can have some important benefits and drawbacks from a legal perspective. When is it legal to record? Once data are recorded, who may access it? Can such recordings be protected from investigative efforts in either civil or criminal legal proceedings? A verbatim record of a suspect's life happenings can serve to either refute or validate allegations. Could personal data archival be the "DNA evidence" of future trials?

1.5 Roadmap

Previous sections have made scattered references to more-detailed descriptions later in the document. This section provides a roadmap.

Chapter 2: The present research connects with many active research areas ranging from classic memory research to contemporary attempts to build similar computer-based memory aids. Chapter 2 pulls the bits out of each one that are most-relevant to the memory prosthesis.

Chapter 3 describes the iRemember memory prosthesis system in detail. This includes the wearable recording device, the server where all data are analyzed, indexed and stored, and all of the memory-retrieval tools available for memory assistance.

Chapter 4 details the evaluations performed on the collected data sets using the tools described in Chapter 3. This includes methods used to capture the data and the methodology used to evaluate the memory-retrieval capabilities.

This work is admittedly an early pass at building a ubiquitous-recording personal memory aid. Chapter 5 organizes the results from the evaluations into design principles for subsequent memory aids. This includes the social and legal ramifications of such ubiquitous recording devices.

Chapter 6: Conclusion

Chapter 2 Related work

Senator, I do not recall

- Lt. Col. Oliver North, during the Iran-Contra congressional hearings

This research touches on a variety of established research areas related to human memory assistance. The first section of this chapter focuses on human memory research relevant to the present work. Included within this is the identification of the memory problems the present memory aid aims to address. Section 2.2 gives a brief survey of memory aids, focusing primarily on computer-based research prototypes. The emphasis of audio in the memory-retrieval approach leads to spoken document retrieval (Section 2.3), which studies how information-retrieval methods can be used on speech. The present research involves the vigilant long-term collection of personal data and experiences. Section 2.4 covers such "pack-rat" practices including the rapidly improving technology to facilitate this.

2.1 Memory

Memory research spans a wide area, and a full treatment of this subject is beyond the scope of this dissertation. This section focuses on research most relevant to computer-based memory aids. It begins with a quick survey of memory research including a taxonomy of memory problems and frequencies that will be used throughout the remainder of the document. Next, it discusses forgetting in real-world situations, expands on the different types of memory failures, and illustrates which ones have the potential benefit of prostheses.

2.1.1 A brief overview of human memory

This section presents a brief overview of material covered in most introductory books on human memory. The three major components of human memory include sensory, short-term, and long-term memory. Sensory memory rests within the individual bodily senses (i.e., sight, hearing, touch, etc.); each receives stimuli and briefly (less than one second) maintains "after images" that are automatically disposed. For that brief time, these sensory inputs are available to the short-term memory system. The level of detail captured by the short-term memory system depends on the amount of attention paid to the sensory inputs. The capacity of short-term memory is sometimes referred to as the "7 ± 2" rule. That is, people can typically retain five to nine elements in short-term memory. Retention in short-term memory is also fleeting (less than one minute), but can be extended through continuous repetition or "rehearsal." Vocal and sub-vocal repetition are common ways of achieving this type of extension. In addition to the notion of short-term memory as a simple, temporary receptacle, recent theories suggest that a more complex "working memory" is integrally involved with the short-term memory system. The working memory system involves cognitive processing and manipulation of short-term information (e.g., adding two numbers in your head). Memories from the short-term or working memory systems can then transition to a more-permanent store called long-term memory. Information stored in long-term memory may last just minutes or could be retained for years. It should be noted that the short-term and working memory systems might call upon long-term memories as part of the cognitive processes involved in short-term manipulation. The iRemember memory prosthesis is concerned with augmenting long-term memory.

The three key processes in the long-term memory system are encoding, storage, and retrieval. Encoding of a new memory typically involves simultaneous retrieval of related memories so that associations and meanings can be assigned. The new memory can then be stored. The strength of the memory (i.e., its resiliency against forgetfulness) depends upon the amount of attention and repetition given to it. The "depth-of-processing" theory suggests that more processing done during encoding leaves a stronger memory trace for future retrieval [5,24]. Such attention and repetition may be spread over multiple sessions spanning long periods of time (e.g., students memorizing information in preparation for an exam).

Long-term memory can be further sub-divided into explicit (declarative) memories and implicit (non-declarative) memories. Implicit memories include procedural skills, conditioned reflexes, emotional conditioning, and priming effects. Explicit memories, the focus of the present work, can be divided into semantic and episodic memories.

Episodic memory, sometimes called "autobiographical memory," refers to memories of specific personally witnessed events. This might include happenings at a social gathering, the name of someone you met, details of a trip to the market, etc. Semantic memory refers to world knowledge, divorced from a specific event, time, or place. For example, the author of "Hamlet" and the meaning of the word "corn" are parts of semantic memory.

There is an interplay between semantic and episodic memory. Information in semantic memory is needed to provide meaning to knowledge in episodic memory. Conversely, certain information in episodic memories can transition to semantic memories. For example, if you go to one professional baseball game, the experience will be part of your episodic memory. However, if you attend many professional baseball games, the details of each episode start to blur into general knowledge of professional baseball games. The present research is mostly interested in explicit memories (i.e., semantic and episodic).

2.1.2 Forgetting

In the late nineteenth century, Herman Ebbinghaus conducted a series of long-term memory experiments quantifying the rate at which people forget [30]. As part of this, he invented the "nonsense syllable" which is a three-letter consonant-vowel-consonant sequence that has no meaning. For example, in English, "GEH" is a nonsense syllable, but "GET" is not. Nonsense syllables are hard to remember, and consequently, well-suited for Ebbinghaus' experiments.

He tested himself by attempting to memorize and recall short sequences of nonsense syllables and varying the time between memorization and recall. These were some of the first controlled laboratory-style experiments studying forgetting and resulted in the seminal "Forgetting Curve" (Figure 2-1). Over subsequent decades, variations on this procedure have been used in laboratory experiments to explore various other dimensions of forgetting (e.g., primacy effects, recency effects, etc.). One of the criticisms and limitations of such experiments is that the environment is sterile and devoid of content and context.

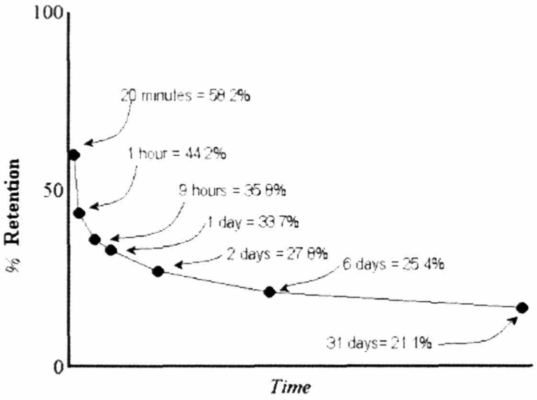


Figure 2-1: Ebbinghaus' Forgetting Curve [30]

Once content and context are involved, we see varying forgetfulness depending upon the different type of memory, the person's age, etc. For example, Warrington and Sanders found that people significantly forget headline news and public events over the years but younger people are more capable of remembering both recent and distant events [118]. Regarding names and portraits of classmates, Bahrick et al. found that even after 30 years, people are good at recognition but are poor at recall [8]. Bahrick and Phelps found that knowledge of foreign languages fades quickly, but is then retained at a constant "permastore" level [9]. Furthermore, these results indicate that a deeper initial learning is still retained even after 50 years (Figure 2-2).

Motor skills are not easily forgotten, but there are differences between "closed-loop" skills that involve continuous physical engagement (e.g., swimming, steering a car) and "open-loop" skills that involve discrete physical motions (e.g., throwing, striking a match); closed-loop skills are more easily forgotten [5,96]. Finally, Conway et al. illustrated how memory retention depends on the type of material being remembered [25].

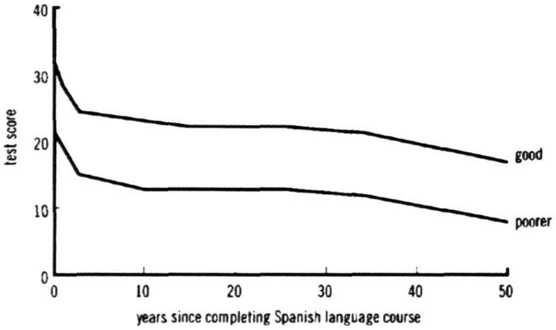


Figure 2-2: Bahrick and Phelps graphs on forgetting of foreign languages. The advantage of better initial learning is maintained over 50 years [9]. Figure from [5].

2.1.3 Schacter's "Seven Deadly Sins of Memory"

Forgetting is not the only type of memory problem, though perhaps the most studied. Schacter outlines the larger space of common long-term memory problems with his "Seven Deadly Sins of Memory" [92] (Figure 2-3).

Forgetting

1. Transience (memory fading over time)
2. Absent-mindedness (shallow processing, forgetting to do things)
3. Blocking (memories temporarily unavailable)

Distortion

4. Misattribution (right memory, wrong source)
5. Suggestibility (implanting memories, leading questions)
6. Bias (distortions and unconscious influences)

7. Persistence (pathological inability to forget)

Figure 2-3: Schacter's "Seven Deadly Sins of Memory" [92]

Transience (Sin #1) can be thought of as the normal degradation of memory accessibility over time. Examples include forgetting facts, details of events, etc. Absent-mindedness problems (Sin #2) can be a result of poor encoding or poor retrieval. Schacter cites examples like forgetting where you left your keys or glasses. A common example of a blocking problem (Sin #3) is the tip-of-the-tongue phenomenon. People suffering from this have the sense that they know the fact they are trying to recall, they might have even remembered it in the recent past, but cannot seem to bring it forward.

Memory distortions illustrate the malleability of human memory. Such effects on memories can be particularly damaging, especially with respect to eyewitness testimony. Misattribution problems (Sin #4) occur when the memory is true except it is attributed to the wrong source; for example, attributing the wrong author to a novel. The sufferer might strongly believe in the truth of their incorrect memory. Suggestibility (Sin #5) allows the incorporation of incorrect information into memory (i.e., "false memories"), often done by deception or leading questioning. Cases of suspects admitting to crimes they did not commit are unfortunate examples of this memory problem [70]. Bias (Sin #6) refers to how current beliefs and opinions influence perceptions of past experiences. The classic study of this showed how couples' present perceptions of their romantic relationships skew their past memories of it [93].

Persistence (Sin #7) can become seriously problematic when it interferes with one's ability to function. For example, those suffering from post-traumatic stress disorder sometimes face unrelenting memories of traumatic experiences.

Throughout the rest of this document, this taxonomy will be used to position pieces of related work and to indicate how the current research fits into the greater body of memory research. The emphasis will be on transience and blocking (Sins 1 and 3) as those are most relevant to the proposed memory aids. This is not to suggest that the computer-based memory aids cannot address the other sins. Rather, as will be argued in the next section, Sins 1 and 3 are simply more common in the context in which the research will be performed. Sin 7 will not be mentioned further as the research aims to help people remember; it is not expected that a memory aid will help people forget.

2.1.4 Frequency of real-world memory failures

Terry performed experiments in non-workplace settings in which subjects maintained "forgetting diaries." Both the frequency and the type of forgetting were recorded. 751 forgetting instances from 50 subjects were recorded and the results suggested that most memory failures are caused by failing to perform some action [107].

Eldridge performed more detailed studies in the workplace with the purpose of designing technological assistance based on common memory problems [31,32]. Through a diary study of 100 people submitting 182 separate memory problems, these memory problems were classified into three categories (Table 2-1):

Problem	Frequency	Description	Example
Retrospective Memory	47%	Remembering past events or information acquired in the past	Forgetting someone's name, a word, an item on a list, a past event.
Prospective Memory	29%	Failure to remember to do something	Forgetting to post a letter, forgetting a lunch appointment
Action slips	24%	Very short-term memory failures that cause problems for the actions currently being carried out	Forgetting to check the motor oil level in the car before leaving on a trip

Table 2-1: Eldridge's classification of common memory problems in the workplace [31]

Under Schacter's taxonomy, retrospective memories are directly analogous to transience. Furthermore, what Eldridge labels "prospective memory" and "action slips" are both forms of what Schacter calls "absent-mindedness." Retrospective memory problems are the most common and out of those, remembering facts and past events form the bulk of the set. This suggests that reminders of past events could help alleviate this problem.

Both Terry's and Eldridge's work found that most memory failures are a result of forgetting to do something instead of forgetting facts. However, Eldridge's work suggests greater fact forgetting occurs in the workplace than in Terry's broad everyday situations. For a memory prosthesis designer, this is fortunate since Eldridge claims that it is easier to design technology to support retrospective memory failures than it is for prospective failures or action slips.

2.1.5 Long-term memory studies

Wagenaar [116] and Linton [68] performed separate, influential multi-year diary studies in which they recorded salient experiences every day into a written journal. For example, for six years, Wagenaar wrote down the time, place, who was with him, and a brief statement about the event. At the end of the recording period, an assistant tested his recollection of randomly selected episodes. The significance of these works rests in their magnitude and application to real-world memories (granted, the retrievals were performed in the laboratory).

Figure 2-4 shows one recall curve from Wagenaar's experiments. This illustrates a sharp decay over the first year and then a steady decay afterward. Figure 2-5 shows retention curves as a function of retrieval cues and suggests more retrieval cues leads to better recall.

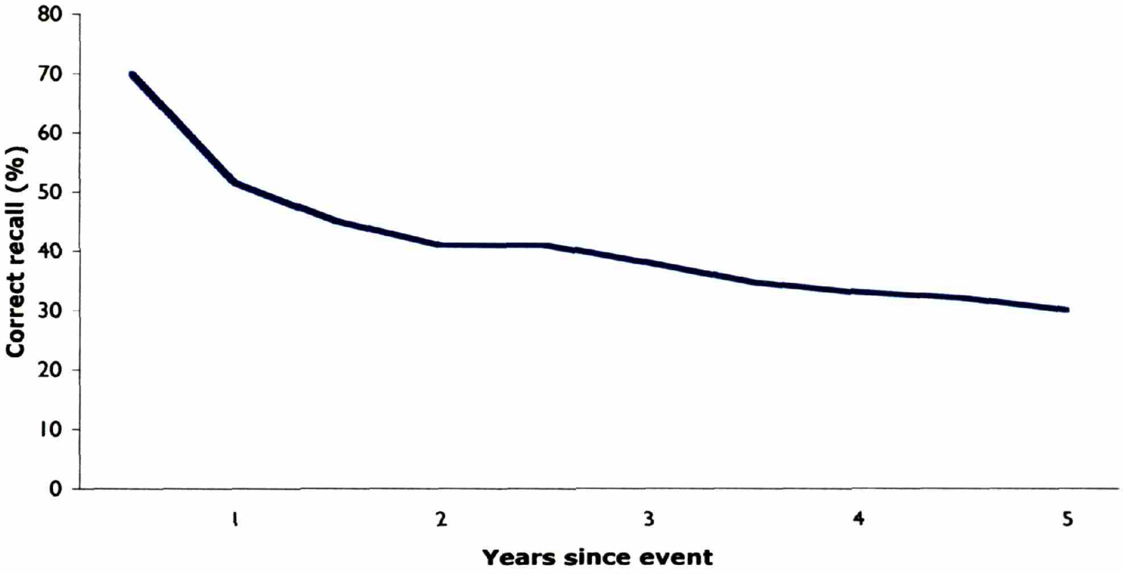


Figure 2-4: Forgetting of salient events [116]

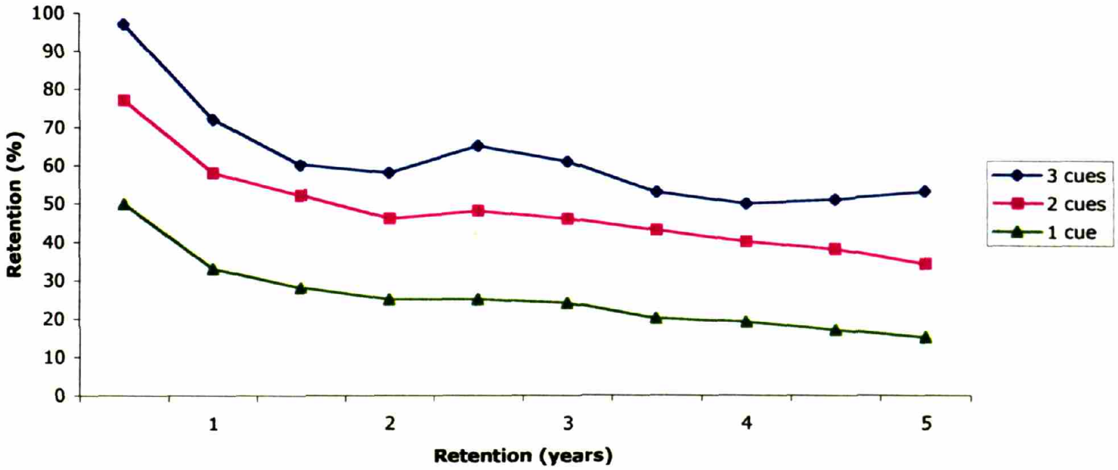


Figure 2-5: Retention curves as a function of the number of retrieval cues [116]

It should be noted that there are differences between Ebbinghaus' forgetting curve, which showed an exponential decay, and Wagenaar's and Linton's results, which both showed linear decay at 5% per year. Linton suggested the difference was due to the salience of real-world experiences versus nonsense syllables.

2.1.6 Landmark events and context dependency

Wagenaar's work suggests that more triggers help elucidate memories. Loftus and Marburger performed experiments demonstrating that some memories are better than others. In particular, "landmark events" are significant episodes either in one's life (birth of a child, graduation) or shared by many (JFK assassination, Space Shuttle Challenger disaster, September 11 terrorist attacks). In the questionnaire study, subjects were asked about crimes committed against them relative to a landmark event that happened six months earlier (the Mt. St. Helens volcanic eruption in northwest United States). They found that subjects could determine the relative timing of the crimes more accurately when put in the context of a landmark event as opposed to exact dates [71]. The problem is called "forward telescoping": people tend to report past experiences occurring more recently than they actually did. For memory-retrieval tasks, having accurate time bounds can be particularly useful in terms of narrowing a list of choices (e.g., search results).

Loftus' and Marburger's work specifically looked at transience memory problems, but one can speculate about the impact landmarks may have on absent-mindedness and blocking. Gooden and Baddeley [44] found that "context dependency"—the ability to remember better when the original context of the desired memory is reconstructed—has an impact on the retrievability of a memory. Intuitively, context dependency suggests one can retrieve a memory better when "returning to the scene of the crime." This phenomenon is most pronounced under physical contexts, but Blaney found evidence suggesting it can be mood-based also [12].

Like Loftus' and Marburger's work, Gooden's and Baddeley's study was limited to transience. To date, there is no research specifically looking at the impact of landmark events and context dependency on absent-mindedness and blocking.

Though some theorize that distinct biological memory-encoding mechanisms exist for landmark events as opposed to "normal" events, there is presently no evidence to support this. Also, different people will have different landmarks based on their individual experiences and what is important to them. Without a means to identify persons-specific salient events, the simplest approach for the memory aid designer to try to capture as many landmark events and as much context as possible.

With regard to contextual dependency, a memory aid with a detailed record of a past event could try to trigger memories by presenting a vivid reconstruction of the event. To a certain extent, that is what the memory-retrieval tools (Chapter 3) provide, albeit in a limited form. The present use of conventional computer displays and interfaces to convey simple textual and iconic representations of the context are likely less-powerful triggers than an immersive virtual-reality or holographic replay. Such interfaces are challenging to construct, mostly in terms collecting the necessary data for such displays.

Contemporarily viable intermediates include photographic and video playback of personal experiences. Again, this approach is beyond the scope of the present work.

2.2 Computer-based memory aids

Memory aids can take different shapes, can be useful for various memory problems, under varying situations, and have varying cost/benefit tradeoffs. Paper and pencil can be a low-cost, low-power, portable memory aid.

This section presents summaries of a sampling of computer-based memory aids. The focus is on research prototypes and most address transience. There are more systems than what is described here and these are meant to cover a sample. The authors of some of these systems did not explicitly label them as memory aids, but the systems have clear memory-assistive value, and are consequently included.

The section is organized based on the approach taken (e.g., text-based, audio-based, etc.) This is not to say that a project fits neatly into the category below or there is not overlap.

2.2.1 Memex

In 1945, Vanevar Bush envisioned a computer-like personal memory aid: memex [18]

> A memex is a device in which an individual stores all of his books, records, and communications, and which is a mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.

Bush never built memex. This is not surprising considering the state of computers in 1945: The ENIAC (one of the earliest computers) was completed in 1945. This was a room-sized monstrosity that required nearly constant maintenance by multiple technicians. Bush's article was meant to be a visionary piece. Even today, the vision has yet to be fully realized, though many (including authors of some the systems cited below) have since cited it as an inspiration.

The remainder of this section will discuss systems that have been implemented.

2.2.2 Active badges

The "Forget-me-not" project [58] included one of the first portable computer-based memory aids: the ParcTab (Figure 2-6). The team coined the phrase "memory prosthesis." The ParcTab is a portable episodic-memory retrieval tool (based on a custom-built handheld computer). It maintains histories of user activity and allows users to search and browse these logs when trying to recall past events. To facilitate recording of user activity, the environment (Rank Xerox Research Center in Cambridge, England) is outfitted with sensors allowing ParcTab to record a limited set of activities including personal location, encounters with others, computer workstation activity, file exchanges, printing, and telephone calls. The interface on the handheld shows a history of past events with iconic representations of people, location and data exchanges.

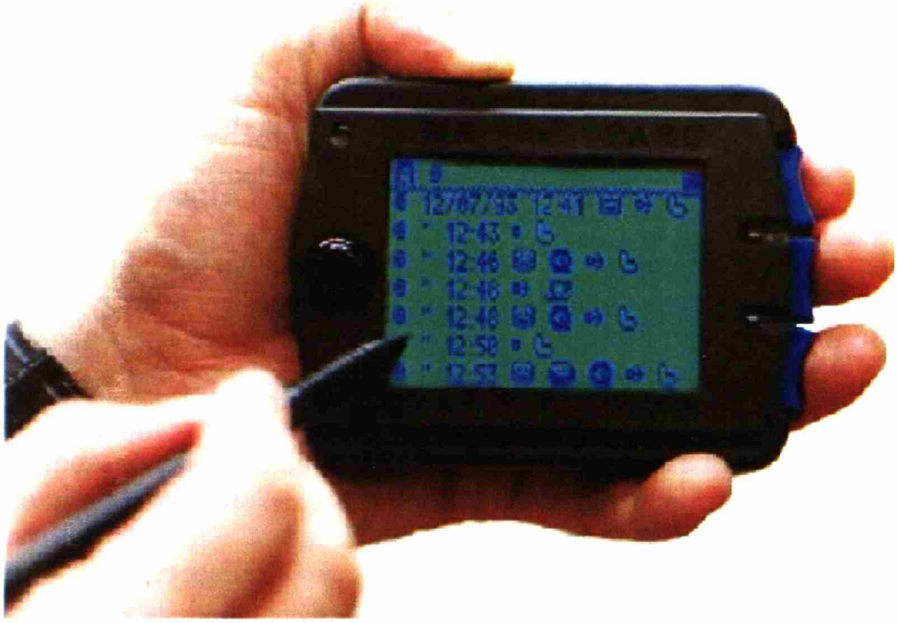


Figure 2-6: The ParcTab "memory prosthesis" used in the Forget-me-not project [58]

The ParcTab was designed to address retrospective (transience) and prospective (absent-mindedness) memory failures. For retrospective memory failures, the ParcTab passively recorded information and allowed the user to review past events when a forgetting incident occured. To assist with prospective failures, it was contextually sensitive to the environment and proactively reminded users of situation-specific information they may need to know or be reminded.

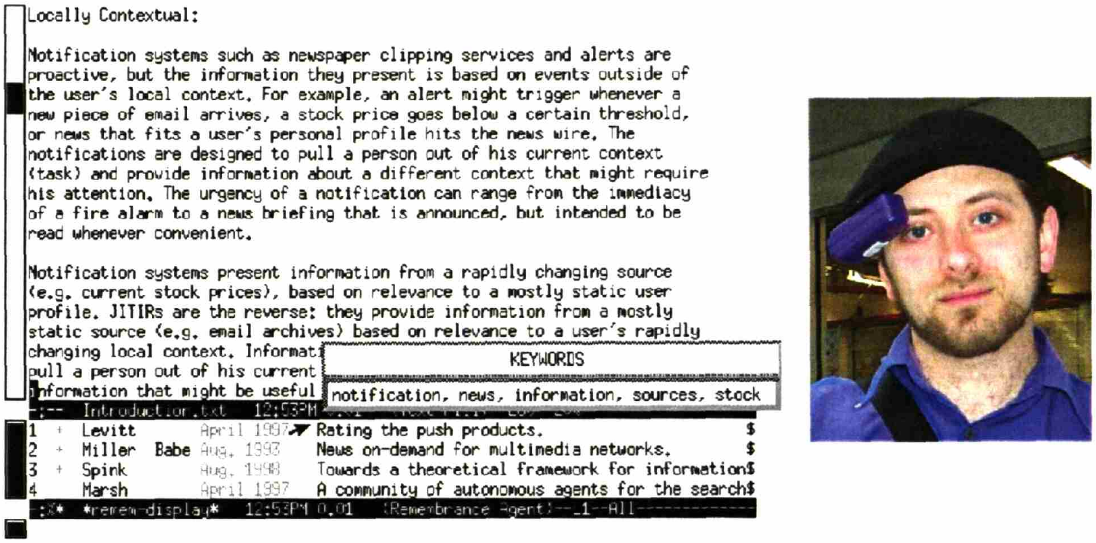
Since ParcTab's data collection was passive, user intervention was not necessary; this was important to minimize the burden of and absent-mindedness in the daily data collection process. ParcTab worked well to log user activity, but it did not store the content of discussions, documents, etc. Consequently, memory failures related to topics discussed, things seen, etc. could not be addressed.

Content-based approaches started to become more popular soon thereafter, enabled in large part by improved storage and wireless networking capabilities of handheld computers in the mid-1990s. Some of these are discussed below.

2.2.3 Text-based approaches

Several projects have approached the memory assistance problem via automatic retrieval of relevant information as a user types into a computer. For example, as I am typing this section, it would be nice if the computer could analyze my text and remind me of research pertaining to computer-based memory aids. Ironically, in an early draft of this section, I forgot to include one important system; fortunately, one of my readers reminded me. Unfortunately, this accidental omission was somewhat embarrassing since the reader was an author on that paper. Had I been using one of the systems cited below, I might have avoided the absent-mindedness.

One system in particular, the Remembrance Agent (RA) allowed users to write notes during conversations in a minimally intrusive way (Figure 2-7) [86]. It was part of the wave of research on wearable computer in the mid- to late-1990s. A one-handed keyboard and a heads-up display allow the user to keep focused on their conversation partner while still writing notes. The computer was carried in a backpack.



```
Locally Contextual:

Notification systems such as newspaper clipping services and alerts are
proactive, but the information they present is based on events outside of
the user's local context. For example, an alert might trigger whenever a
new piece of email arrives, a stock price goes below a certain threshold,
or news that fits a user's personal profile hits the news wire. The
notifications are designed to pull a person out of his current context
(task) and provide information about a different context that might require
his attention. The urgency of a notification can range from the immediacy
of a fire alarm to a news briefing that is announced, but intended to be
read whenever convenient.

Notification systems present information from a rapidly changing source
(e.g. current stock prices), based on relevance to a mostly static user
profile. JITIRs are the reverse: they provide information from a mostly
static source (e.g. email archives) based on relevance to a user's rapidly
changing local context. Informati
pull a person out of his current
information that might be useful
-:-- Introduction.txt     12:53PM
1  +  Levitt        April 1997   Rating the push products.                     $
2  +  Miller  Babe  Aug. 1993    News on-demand for multimedia networks.       $
3  +  Spink         Aug. 1998    Towards a theoretical framework for information$
4     Marsh         April 1997   A community of autonomous agents for the search$
-:%*  *remem-display*   12:53PM 0.01   (Remembrance Agent)--_1--All------------
```

Figure 2-7: The Remembrance Agent user interface (left). This is what is visible through the ocular piece of the wearable. Photo of Rhodes wearing the Remembrance Agent (right) [86].

While the user is typing, the RA proactively searches through all previously written notes and identifies the most-relevant ones. This identification is done using the most-recently typed words as a query to a relevance-ranking keyword search engine; the results are presented in the same head-up display. Since the search process is fully automated, the user can at-a-glance see if any past notes are relevant to the just-written note. Watson [17] and Reflecting Pool [13] employed similar techniques, but were targeted towards document editing and note taking on desktop- and laptop-computing scenarios.

In contrast to the ParcTab, the RA did not passively capture data and was dependent on the user's data entry vigilance. Conversely, the RA, by virtue of allowing unconstrained text entry, permitted the inclusion of any text, including conversation content: something the ParcTab did not allow.

Rhodes collected several years worth of notes and cited anecdotal evidence suggesting the RA helped mainly with transience and absent-mindedness, but also helped with misattribution, suggestibility, and bias to a lesser degree [Rhodes pers. comm. 2001].

2.2.4 Audio-based systems

The text-based approaches described earlier require that users document events in their lives in order to get some future benefit. This may be too much to ask for many people. But, even assuming one who is dedicated and motivated to document their lives (e.g., the principal investigator on the research project), most people's typing rate is slower than the average speech rate (180 words per minute); relevance decisions must be made on the spot; some data will not be captured.

Audio recording can help with the data capture process since a typical audio recorder requires much less effort to capture conversational data: turning it on and turning it off. If the conversation is long and one is using a tape-based audio recorder, one might need to swap tapes. The appeal of this approach to capture data is not without a downside. When it comes time for retrieval, content-based search of audio is much harder than text. This is mainly due to the error-prone nature of large-vocabulary automatic speech recognition. This topic, better known as "spoken document retrieval," will be discussed in detail in Section 2.3. The remainder of this section describes systems that take alternate approaches to audio retrieval.

Many of the non-speech-recognition audio-retrieval approaches in the 1990s relied on alternate audio analysis technique, visual representations, a user-based marking scheme, or some combination thereof [2,26,52,77]. At that time, most attention was paid to telephone and meeting room settings since it was easier to integrate computers into these environments.

These approaches took a leap forward with projects like the Audio Notebook [106] (Figure 2-8) and Filochat [124]. These prototypes were designed for note-taking situations (e.g., lectures, meetings, etc.). They allowed users to make handwritten notes on a digital notepad while simultaneously audio recording. The audio is automatically indexed based on the handwritten stroke occurring at the same moment. Selecting a section of handwritten text retrieves the corresponding audio segment recorded at the time of the handwritten stroke. Both the audio recording and written record serve as memory triggers to the past event. Given the context under which both projects anticipate use, both are optimally assistive to transience memory failure.

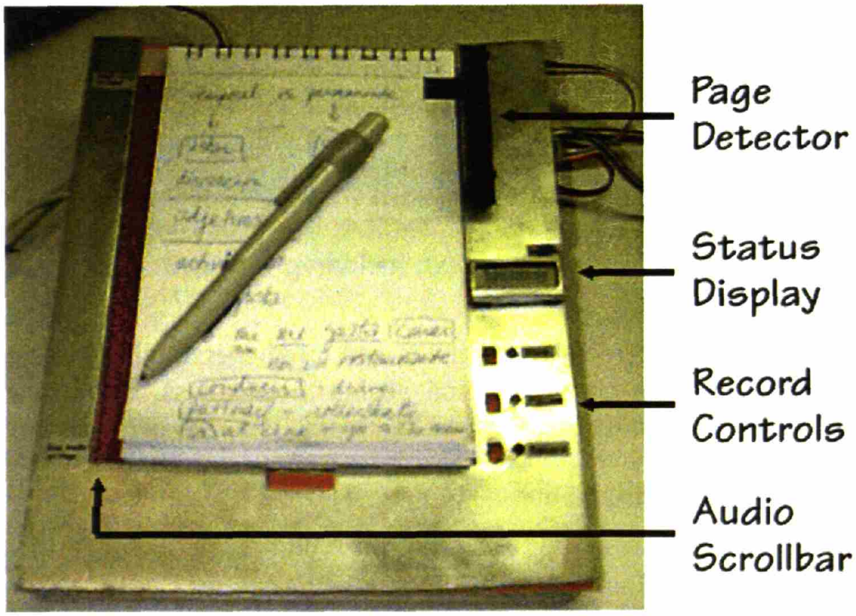


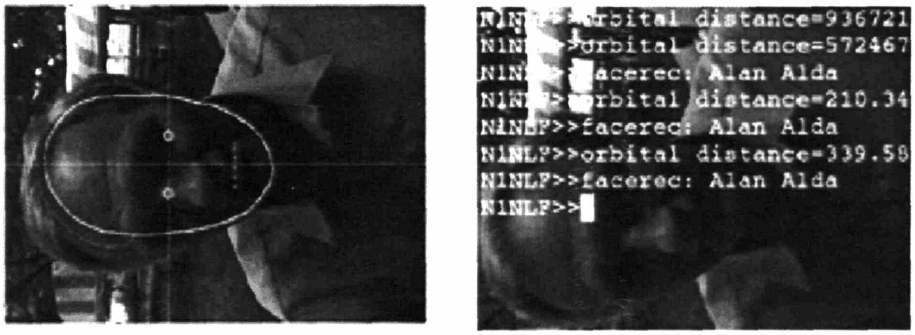
Figure 2-8: The Audio Notebook prototype [106]

Another audio-based system, ScanMail [122], was designed for voicemail retrieval and will be discussed in more detail in the context of spoken document retrieval (Section 2.3.2). Another contemporary system, the Personal Audio Loop [47], is a portable recorder that captures verbatim audio for the purpose of memory assistance but is targeted towards remembering information within the past day. In fact, its data storage automatically purges audio after a day. Retrieval depends more upon the user's remembrance of events in the past day to help localize audio searches.

2.2.5 Photo- and video-based systems

Video-based approaches have historically been hampered by poor portability of video cameras and limited video analysis techniques. Like audio, this too is changing rapidly enabled in large part by miniaturization, advances wearable computers, and improvements in content-based video retrieval [102].

WearCam [74] was designed as a "visual memory prosthetic (sic)" with the goal of alleviating "visual amnesia." The system included a wearable camera, a facial recognition system, and the ability to provide "flashbacks" or visual replays of recent images. These replays allowed the wearer to better encode their experiences by being reminded of previously experienced faces. Figure 2-9 shows images from WearCam's facial recognition system.



```
N1N    rbital distance=936721
N1N   >orbital distance=572467
N1N     acerec: Alan Alda
N1N   > rbital distance=210.34
N1NLF>>facerec: Alan Alda
N1NLF>>orbital distance=339.58
N1NLF>>facerec: Alan Alda
N1NLF>>
```

Figure 2-9: WearCam. Left image shows template and right image shows information about match. Images copied from [74]

StartleCam [48] also includes a wearable apparatus to capture video sequences from daily life, but selectively stored sequences based on identification of biometrically identified startle response. SenseCam [127] is a similar, more-recent effort.

2.2.6 Temporal- and landmark-based systems

The previous few sections have discussed retrieval approaches based on various types of content. This and the next section talk about alternate data organizational schemes that further leverage one's memory in the retrieval process.

Temporal organization allows the use of landmark events to localize data searches. For example, users could use more-memorable landmark events to help find less-memorable events. Studies have shown benefits of landmarks with computer-based retrieval of personal information [87].

MEMOIRS [59] and Lifestreams [38] are user-interface systems designed as alternatives to the "desktop metaphor" commonly found on personal computers. The consistent idea between the two—relevant to memory aids—is that users' document interactions (i.e., creation, editing, etc.) are automatically organized as a stream of temporal events. Instead of navigating hierarchically organized folders, users may browse temporally using their episodic memory of past document-manipulations. Stuff I've Seen [28] offers a similar episodic, temporal organization and includes email, calendar, web-browsing activity, as well as document activity. All three of these systems are targeted towards transience, but the detailed histories they collect can also help with any of the first six memory sins.

2.2.7 Spatial organization

The previous three projects encourage use of temporal memory triggers for information navigation; the next two advocate physical organization. The Spatial Hypertext project suggests that people remember spatial arrangements of information better than navigation paths [98]. The contextual dependency theory gives credence to this approach.

Piles [89] uses a metaphor of organizing information into stacks (similar to piles of paper on a desk) instead of hierarchical folders. Piles tend to have less structure than folders or directories and do not force users to commit to premature partitioning and organization of their documents. Users can arrange these piles in physical positions on the computer screen. To help users understand their partitioning schemes, Piles examines the contents of all documents in a given pile, determines similarities, and presents these to the user when they browse the pile.

2.3 Spoken document retrieval

When faced with a transience or blocking memory problem, one may remember a variety of cues to help formulate a search strategy. Some cues may be contextual (e.g., where, when, who, etc.). Such data can be gathered with reasonable accuracy using commercially available solutions such as a mobile phone enabled with global positioning system and Bluetooth® readers. Others cues may pertain directly to the content (e.g., something that was said). If so, the ability to quickly find the conversation or passage within collections of recorded speech would enhance a person's ability to formulate search strategies based on remembered content.

The goal of spoken document retrieval research is to build systems that can locate words or phrases within collections of audio recordings. This has been studied in detail as part of the Text Retrieval Conference (TREC) Spoken Document Retrieval (SDR) track [41]. The typical approach is to use a large-vocabulary automatic speech recognizer (ASR) [84] to transform audio into a text transcript that, in turn, can be searched using conventional keyword-based information-retrieval techniques [91]. This deceptively simple statement does not reflect the true complexity and challenges of the task.

Two separate, yet related problems in text-only information retrieval are particularly exacerbated in speech archives: (1) finding the correct "document" within a collection; and (2) localizing within a "document" for the desired information. With respect to (1), transcription errors lead to both false-positive and false-negative search errors that are manifested to users as poor precision, recall, and relevance-rank ordering. TREC SDR track has focused on audio from the broadcast newscasts. Error rates are much higher for spontaneous speech [117], the domain of present interest. Unfortunately, customized ranking algorithms for error-prone ASR-generated transcripts are no better than those used for low-error written text [99]. With regard to (2), Whittaker et al. suggest that the larger problem in searching speech archives is browsing and localization difficulties within an individual speech document [123].

In search tasks, visualizations of speech-recognizer-generated transcripts and spoken document retrieval serve as an alternate to detailed re-listening of audio. Traditional SDR studies information-retrieval techniques with the assumption that someone is searching through an unfamiliar collection. With memory retrieval, one need not make such an assumption and can expect subjects' remembrance of past events to help. As stated in Chapter 1 differences are expected among the search strategies employed by those trying to remember a previously witnessed event versus someone trying to find information within an unfamiliar collection; in the former case, retrieving any information that triggers the memory of the event is sufficient; in the later, finding the exact information is necessary.

2.3.1 Speech recognition errors

Speech-recognizer-generated transcripts almost always suffer from poor recognition accuracy, are difficult to read, can confound readers, and can waste their time [122]. Successful speech recognizers tend to achieve better accuracy when limiting vocabularies, speakers, speech style, and acoustic conditions. When measuring speech-recognition accuracy, the most-common metric, word error rate (WER), is defined as follows:

$$\text{WER} = \frac{\text{insertions} + \text{deletions} + \text{substitutions}}{\text{words in perfect transcript}}$$

Most errors stem from several sources. First, sub-optimal acoustic conditions can increase WER. Common causes include low-quality microphones, sub-optimal microphone positioning relative to the speaker, and excessive background noise. Second, out-of-vocabulary (OOV) errors occur when a speaker says a word that is not included in the recognizer's vocabulary. When faced with an OOV word, a speech recognizer will typically output an erroneous word or words that may sound like the desired word.

To lower WER, one can use better microphones (e.g., noise-canceling) or position the microphones better (e.g., closer to the speaker). Most large-vocabulary speech-recognition systems offer a procedure called "enrollment" in which the recognizer requests voice samples from a speaker and adapts to the characteristics of that speaker's voice. One type of enrollment called "cepstral normalization" [69] is a procedure whereby a speech recognizer adapts to the acoustic characteristics of a particular microphone.

Other ways to reduce errors include limiting the vocabulary and mandating users' speech conforms to formulaic grammatical patterns (e.g., a context-free grammar). Not every application can operate under these constraints, but the contemporary speech-recognizer-based systems that are achieving some level of success are typically constrained in this manner (e.g., speech-based telephone menu navigation, airline information systems, etc.).

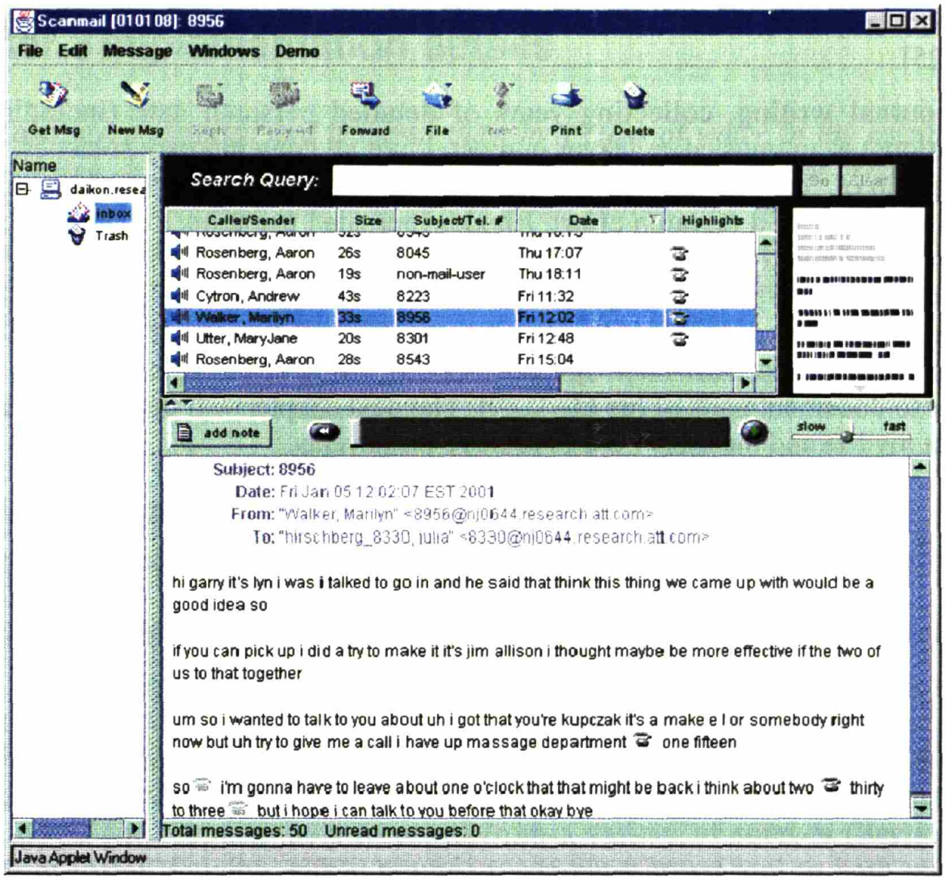
Optimistic predictions aside, high-accuracy speech recognition of conversations in poorly microphoned, heterogeneous environments will not happen anytime soon. Despite this, high-quality transcription—while beneficial—may not be necessary, especially for memory-retrieval tasks. Witbrock [128] suggests general-purpose, spoken-document retrieval tasks can still be performed at high WER. Speech recognition has been shown to help in voicemail-retrieval [122] and calendar-scheduling tasks [129]. Section 4.3 introduces the "speed listening" techinique. This illustrates how error-laden speech-recognizer-generated transcripts synchronized with time-compressed audio playback can improve subject comprehension, especially when word-brightness is rendered proportional to recognizer-reported confidence.

Although we are still far from the panacea of high-accuracy, speaker-independent, large-vocabulary recognition systems that would enable a vast array of speech applications, the state of speech recognition is nearing the point in which a limited set of new applications would benefit from speech recognition even with the limited accuracy found in today's recognition systems. The next section describes one such application.

2.3.2 ScanMail

An example spoken document retrieval system is ScanMail [122]. It provides a graphical user interface similar to an email client, but for voicemail (Figure 2-10). The system also includes speaker identification technology and uses large-vocabulary automatic speech recognition technology (ASR) to transcribe the voicemail. ScanMail's design is based on a set of studies that suggest frequent voicemail users tend to use such systems for tracking action items and other pending tasks [123]. This suggests that in practice it could be used to alleviate absent-mindedness. Presently, no formal studies have been conducted to examine this point.

ScanMail also represents one of the only working systems designed to use ASR for browsing, searching, and retrieving past events—in this case, voicemail. However, the studies of ScanMail are inconclusive with respect to the value of ASR in that system. The prototype memory-retrieval tool (Section 3.5) designed for the evaluations was influenced by the ScanMail design and expands upon it to accommodate the large quantity of data accrued from conversational data (as compared to voicemail messages).



Caller/Sender	Size	Subject/Tel. #	Date	Highlights
Rosenberg, Aaron	26s	8045	Thu 17:07	☎
Rosenberg, Aaron	19s	non-mail-user	Thu 18:11	☎
Cytron, Andrew	43s	8223	Fri 11:32	☎
Walker, Marilyn	33s	8956	Fri 12:02	☎
Utter, MaryJane	20s	8301	Fri 12:48	☎
Rosenberg, Aaron	28s	8543	Fri 15:04	

Subject: 8956
Date: Fri Jan 05 12:02:07 EST 2001
From: "Walker, Marilyn" <8956@nj0644.research.att.com>
To: "hirschberg_8330, julia" <8330@nj0644.research.att.com>

hi garry it's lyn i was i talked to go in and he said that think this thing we came up with would be a good idea so

if you can pick up i did a try to make it it's jim allison i thought maybe be more effective if the two of us to that together

um so i wanted to talk to you about uh i got that you're kupczak it's a make e l or somebody right now but uh try to give me a call i have up massage department ☎ one fifteen

so ☎ i'm gonna have to leave about one o'clock that that might be back i think about two ☎ thirty to three ☎ but i hope i can talk to you before that okay bye

Total messages: 50 Unread messages: 0

Figure 2-10: ScanMail interface [122]

2.4 Personal data archival

Collecting an archive of personal experiences is an important step towards building an aid that can help remedy long-term real-world memory problems. Diligent archival of such data is not unheard of. Ask anyone who regularly maintains a diary or a journal. The "Remembrance Agent" (Section 2.2.3) could be used for journal writing and illustrated some of the value of doing so on a computer and *in situ*.

In the past few years, there has been an increased interest in systems that expand on this idea. These range from computer-based tools that augment the construction of personal diaries, weblogs, etc. by simplifying data entry [64] to general purpose capture systems intended to accrue every bit of data a person experiences throughout their lives [42]. Again, this is enabled in large part by the miniaturization of digital recording devices and the low cost of digital storage.

Archival of personal data is not new. Long-term archival of files on personal computers hard disks and email is already commonplace. Often, important data are not only stored, but are duplicated as backups. For certain data, automatic archival is typically the default behavior (i.e., email systems typically do not delete messages once read) and space-savings has become a less-prominent reason for deletion. In fact, it could be argued that it is easier to keep data than to delete. Management of email stores has been studied in detail [125].

Unlike journal writing, collecting years of detailed personal data (including daily conversations) is an unusual endeavor in most, if not all contemporary societies. There is a small group of "diligent archivers" who have started along this path. Not surprisingly, most do so as part of a research effort. Rhodes journaled using his Remembrance Agent [86]; Orwant captured computer activity with doppelgänger [83]; Gerasimov captured his biometric signals [43]; both Wagenaar and Linton documented daily events as part of their memory experiments [67,116]; Bell is currently archiving all of his documents and electronic communications with MyLifeBits [42]; Wong at Georgia Tech also wears a computer to document parts of his life [129]. The next chapter will describe what I use to document my life and the tools I use to retrieve past events.

Chapter 3 Implementation details

> *Take seven almonds and immerse them in a glass of water in the evening. Next morning, after removing the red skin, grind the almonds. Mix the ground almonds with a glass of milk and boil. When it has boiled, mix in a spoonful of* ghee *(clarified butter) and two spoonfuls of sugar. When it is lukewarm, drink it. Take this on an empty stomach in the morning and do not eat anything for the next two hours. Do this for 15 to 40 days.*
>
> - Indian home remedy for "weak memory"

3.1 Design goals

This section first spells out the design goals of the technology before describing the iRemember system in detail. Some of these topics have been discussed earlier, but it is worthwhile highlighting some of the relevant points and refreshing the reader's memory.

Section 2.1.3 described Schacter's "Seven Sins of Memory." It is not the goal of the present research nor is it expected that a single memory aid can adequately address all such problems; instead, the focus is to address a subset. Specifically, the prototype aims to address transience, the most frequent among the memory problems. The approach is to collect, index, and organize data recorded from a variety of sources (Section 1.2.4) related to everyday activity, and to provide a computer-based tool to both search and browse the collection. The hope is that some fragment of recorded data can act as a trigger for a forgotten memory.

It is anticipated that blocking problems would also benefit from such an aid. One of the common qualities of both transience and blocking is that the person is aware of the memory problem when it occurs (this is not true for all memory problems). Assuming the person also wishes or needs to remedy the problem, what is needed is a resource to help. This is where the memory prosthesis comes into play.

As stated in Chapter 1, the data capture approach advocates nearly passive audio recording encompassing many daily-life situations via computer-based portable recording devices. The specific device will be introduced in Section 3.3.1. The choice to prefer passive data capture is to simplify and reduce forgetfulness in the daily data-capture process. Ironically, forgetting to activate the memory prosthesis (i.e., absent-mindedness) is a common problem.

Given the present focus on the workplace settings, it is anticipated that users could benefit from tools on both handheld/wearable devices and desktop/laptop personal computers. Hence, implementations were made for all of these platforms, taking advantage of the idiosyncrasies of each. It should be noted that evaluations in Chapter 4 focus on the wearable capture device (Section 3.3.1) and the personal-computer-based retrieval tool (Section 3.5). Consequently, the emphasis in this chapter will be on those tools.

3.2 System overview

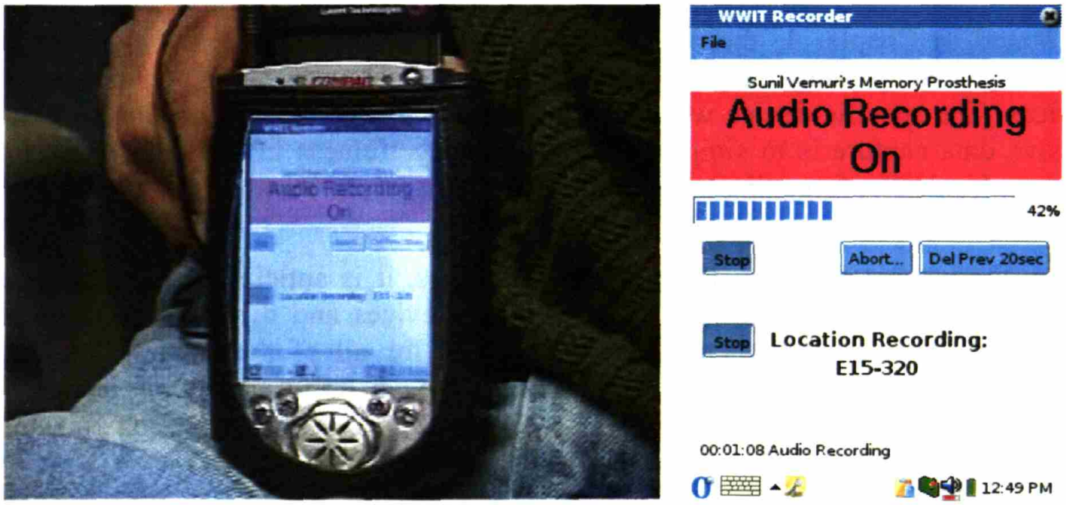
The iRemember system is fairly straightforward: a set of recording apparatuses and retrieval tools. Between all of these is a high-capacity server that receives all data from the recording apparatuses, processes the data, and makes it available the retrieval tools.

The tools serve as a platform for experimentation. The current implementation has limitations and similar devices and software in wide-scale use by a general populace would have properties akin to a commercially developed product (e.g., longer battery life, smaller form-factor, better user-interface and industrial design, more-stable software, etc.). The current goal was to allow users who are sympathetic to the research to start experiencing portable, ubiquitous recording and participate in the evaluations described in Chapter 4.

3.3 Recording tools

3.3.1 Wearable recording apparatus

The wearable prototype (Figure 3-1) is designed to be a nearly always-available data-capture tool. It is an iPaq 3650 handheld "personal digital assistant" (PDA) coupled with an 802.11b wireless network card. When fully configured, the device's dimensions are 15.5cm(h) x 8.5cm(w) x 3.1cm(d) and weighs 350 grams (12.3 ounces). This was one of the original iPaq models dating to early 2001. At that time, different hardware platforms were evaluated for this project and this iPaq was considered one of the premiere PDAs on the market at any cost. The consumer electronics marketplace is a highly competitive that often experiences rapid engineering improvements; it should be no surprise that presently available iPaqs are smaller, weigh less, cost less, and are superior in other ways (faster CPU, more memory, etc.). Also, other devices have emerged on the market (e.g., programmable mobile phones, high-capacity voice recorders) that would also be viable wearable recording apparatuses. However, these alternates are not discussed and the prototype running on the iPaq 3650 was used for the duration of the project.



Sunil Vemuri's Memory Prosthesis

Audio Recording On

Location Recording: E15-320

00:01:08 Audio Recording

Figure 3-1: Prototype "memory prosthesis" (left) with close-up of screen on the right

The device runs the Familiar Linux [36] operating system with the Open Palmtop Integrated Environment (OPIE) graphical user environment [81]. Familiar Linux a derivative of the Linux operating system customized for handheld devices and OPIE derives from the Qtopia environment. Both are open-source development projects. Custom data-recording software was written on this platform. A screenshot of this can be seen in Figure 3-1. The user may independently activate audio and physical location recording.

The device determines location via proximity to numerous stationary 802.11b base stations primarily inside the building. As opposed to 802.11b "triangulation" methods [7,20,75] that determine location using the combined signal strengths from multiple, nearby base stations, the prototype determines location using only a single, "primary" base station. This primary-base-station-only approach combined with the distribution of base stations in the Media Lab gives horizontal accuracy within roughly 40 feet on a given floor. This method, for the most part, narrows the location to a limited set of lab areas and correctly identifies the floor the device is located. However, the Media Lab's wireless network was designed such that some base stations service areas on multiple floors. Vertical location accuracy suffers in these areas (i.e. wrong floor). The triangulation method, while more accurate (roughly ten feet), requires significant time for initial calibration of the environment and periodic re-calibration. The primary base station approach, with its lower maintenance benefit, was deemed satisfactory

Audio is recorded digitally (22050 samples per second, 16-bits per sample, mono) via either the iPaq's built-in near-field microphone or an external lavalier microphone (Audio-Technica ATR35s). Digitally recording at lower sample rates and sizes results in significantly higher speech-recognition errors. Recording at the aforementioned sample rate and size requires roughly 2.5 megabytes per minute of storage. Under these parameters, the limited storage capacity (32 megabytes) of the iPaq 3650 holds just over 12 minutes of audio: this is shorter than desired. Hence, data are streamed in real-time to a high-capacity server. Part of the reason the iPaq was chosen was because, during early development, it was one of the smallest, portable devices capable of speech-recognition-quality audio recording *and* wireless computer networking. Unfortunately, one of the consequences of activating wireless networking on the iPaq is the heavy drain on the battery. The device usually lasts between 60 and 90 minutes before requiring a recharge.

The real-time data streaming requirement limits use of the device to areas where an 802.11b wireless network is available. To accommodate the data-security requests of the users, further restrictions were imposed limiting data streaming to within the Media Lab's local computer network. Prior to these requests, the device was used to record in settings outside the local computer network, including nearby publicly accessible wireless networks and my apartment. The server receives and stores all data; it also passively captures the additional sources mentioned in Section 1.2.4 (e.g., weather and news-related web sites).

The design reflects a variety of considerations necessary to balance the requirements of obtaining speech-recognition-quality audio while meeting the social, legal, and human-subjects requirements. Moreover, the investigators tried to satisfy a higher standard by ensuring co-workers felt reasonably comfortable with the presence of potentially privacy-invading technology. The device includes a display with the audio recording status, an audio-level meter, and recordees may request recording deletion via anonymized email.

When recording, I often move the device from its waist-level belt-clip in favor of chest height. This allows conversation partners a clear view of the screen, the recording status of the device, and serves as a constant reminder that the conversation is being recorded. Chest-height positioning also provides better acoustic conditions for the built-in near-field microphone on the iPaq. The external, attachable, lavalier microphone is also occasionally used. Although speech-recognition-quality recording of all participants in a conversation is desired, this is not always feasible since no cost-effective, portable recording solutions could be found. Placing the iPaq halfway between two speakers results in bad audio for both. Placing the device at chest-height or using the lavalier allows at least one of the speakers, the wearer, to be recorded at adequate quality at the cost of poor-quality recordings for all other speakers.

3.3.2 Personal computer recording apparatus

Handheld PDAs are one form of portable computers. Another common, portable computer is a laptop. Though physically larger than a PDA, they are just as capable if not more of performing the necessary recording tasks. Moreover, if users are already in the habit of carrying a laptop computer, asking them to tote the iPaq for the purpose of audio recording is redundant. To accommodate such users, iRemember includes a simple audio recording application that can be run on a laptop computer.

Since laptop computers often have large-capacity hard disks (unlike the iPaq), the software does not stream audio recordings in real-time to the server. This has the added benefit of allowing users to record even when there is no wireless network connection available. When the user has connected the laptop back to the Media Lab's computer network, they can submit the recording to the server for inclusion in their data set.

3.4 Server-side data processing

An off-the-shelf version of IBM's ViaVoice® speech-recognition software [114] is used to convert recorded speech to text. Along with each word, ViaVoice® reports the "phrase score" and start and stop time offsets when the word was uttered in the recording. "Phrase score" is documented as follows: "[it] is not a confidence… it is an average acoustic score per second. The acoustic score depends on the quality of the match and the length of the speech aligned with the word" [115]. To better understand the meaning of phrase score in relation to the speech recordings used in the evaluations, some conference talks were recorded and hand-transcribed. These transcripts were then compared with the speech-recognizer-generated ones. Figure 3-2 illustrates the correlation between phrase score and recognition rate (r_s=0.9385, p<0.0001).

Section 3.5.4 introduces the interface for displaying speech-recognizer-generated transcripts. The phrase score is used to vary brightness of individual words based on recognizer-reported confidence. The start- and stop-times of each word are used to determine paragraph and boundaries.

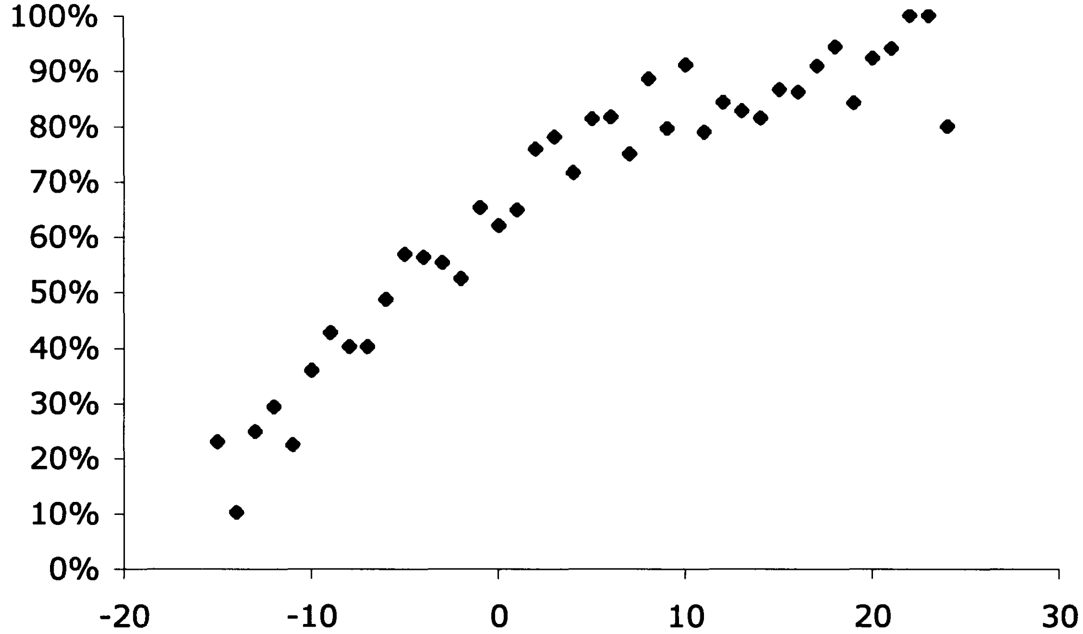


Figure 3-2: Percent of words recognized correctly at each recognizer-assigned "phrase score" (~2,300 words, minimum 10 words per score)

The audio received by the server is in the original 22050 Hz, 16-bit, mono, uncompressed pulse-code-modulation format (more commonly known as uncompressed "wav"). Speech recognition is performed on these uncompressed recordings. After speech recognition is complete, the wav-formatted files are compressed to MPEG 1 Layer III format (more commonly known as "mp3") at a fixed data rate of 32 kilobits per second (Kbs) using the Lame encoder [56]. This is meant to both save space and reduce the data transmission time when users want to download audio recordings from the server for playback via the retrieval software. At 32 Kbs, one minute of audio occupies 240 kilobytes (compared to the original recording which takes up roughly 2.5 megabytes per minute). The compression ratio is slightly better than 10.5:1. With mp3 audio compression, audio quality degrades as data rate decreases (as judged by human listeners). Based on listening to samples recorded by the iPaq, I determined that 32 Kbs was the lowest data rate (i.e., highest compression) that maintained intelligibility of the audio on the two targeted playback devices: the iPaq and a laptop computer. There are alternate audio-compression schemes that claim higher quality audio at lower data rates (e.g., Advanced Audio Coding [AAC], Qualcomm Purevoice™, etc.). However, all of these are proprietary or are available on limited computer platforms. Software that can play mp3-format audio are free, widely available on all desired computer platforms, and are available as software libraries that can be included as part of all implementations of the memory-retrieval tool.

3.5 Memory retrieval tools (personal computer)

An anticipated consequence of a daily-worn recording device is the accrual of years of personal interaction and conversation data. To help alleviate transience and blocking memory problems, it is expected that users would want to search such a collection for memory triggers. Given the focus on audio and the state-of-the-art in information retrieval technologies, capabilities such as browsing, skimming, and free-form keyword searches of timeline-organized audio archives were selected for implementation.

Details of the specific implementations of these are presented in this section. Refining these involved an iterative process with intermediate evaluations. Some of these evaluations were small-scale pilot studies that are not reported in this dissertation. The larger efforts are reported in Chapter 4. One validates the memory-retrieval tool using data collected from conference talks (Section 4.2). Another evaluates the "speed listening" technique developed as part of this dissertation (Section 4.3). Speed listening is not about memory per se, but helps reduce the time needed to listen to audio. A third evaluation evaluates the memory retrieval approach using several years of conversations I recorded with iRemember (Section 4.4).

Memory-retrieval tools were implemented on two platforms. The first runs on a desktop or laptop personal computer. The second runs on the handheld PDA. In a contemporary "knowledge worker" environment, most people have such tools. On a more practical level, the ease of implementing features (some of which are computationally and graphically demanding) compared to a handheld PDA makes the personal computer the preferred platform for the first prototype. These tools will be described in detail later in this chapter.

Little is known about memory retrieval using personal data archives. Hence, one goal of the evaluation is to begin understanding how users approach memory repair and which among the many features prove worthy of inclusion in future iterations.

3.5.1 Searching audio collections

Figure 3-3 shows a screenshot of the personal-computer-based interface available to search and browse large collections of recordings. The top section shows a multi-year timeline of all data. A short, colored vertical tick mark, represents each item. The color corresponds to one of audio (yellow), calendar (blue), email (green), weather (red), and news (orange). Users can selectively show or hide each data type. In the figure, only audio and calendar data are shown.

Below the timeline is a zoomed-in view of a region. In the figure, one week is shown. Each column along the x-axis corresponds to a single day; time-of-day is on the y-axis. The visual allows the simultaneous display of several data types: audio recordings, calendar events, email, and weather. For clarity, only a subset of the data is shown in the screenshot.

Each calendar event has a corresponding blue rectangle. Audio recordings are represented with a yellow rectangle and an associated icon 🔊. Zooming and panning features are provided; double-clicking on an item opens a more detailed view of a recording. Color transparency is used to blend overlapping elements. Cluttering of text from neighboring items is possible and individual items can be made more legible by clicking on them.

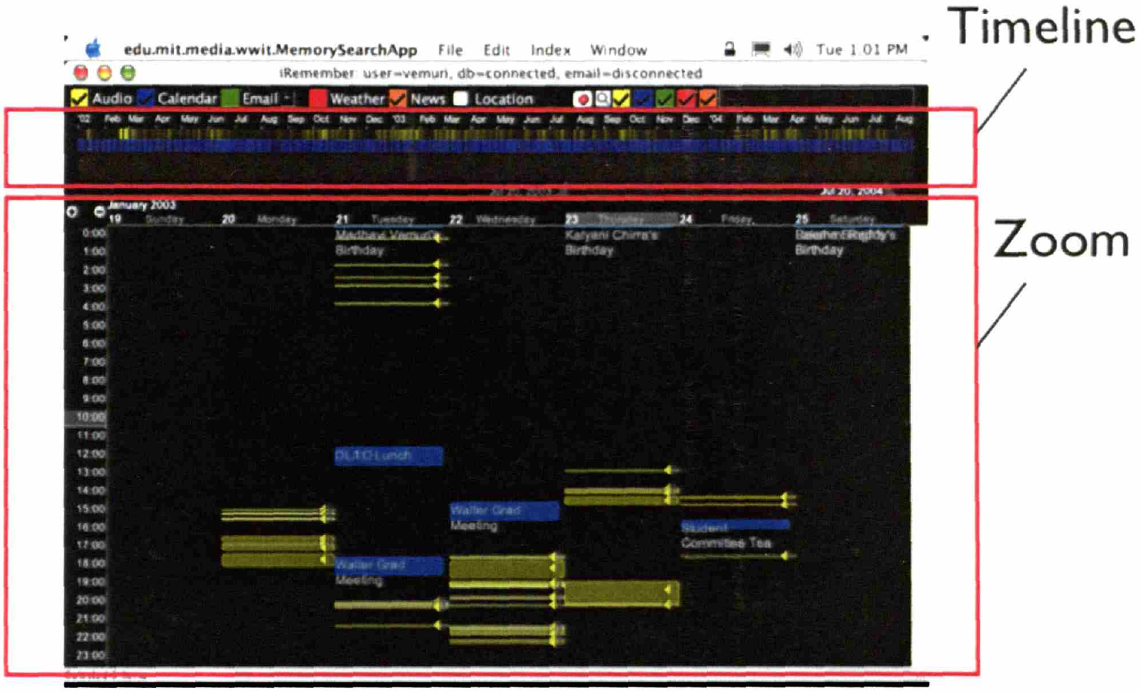

edu.mit.media.wwit.MemorySearchApp
File
Edit
Index
Window
Tue 1:01 PM
iRemember: user=vemuri, db=connected, email=disconnected
Audio
Calendar
Email
Weather
News
Location
'02 Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec '03 Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec '04 Feb Mar Apr May Jun Jul Aug
Jul 20, 2004
January 2003
19 Sunday
20 Monday
21 Tuesday
22 Wednesday
23 Thursday
24 Friday
25 Saturday
Kalyani Chirra's Birthday
Birthday
Walter Grad Meeting
Timeline
Zoom

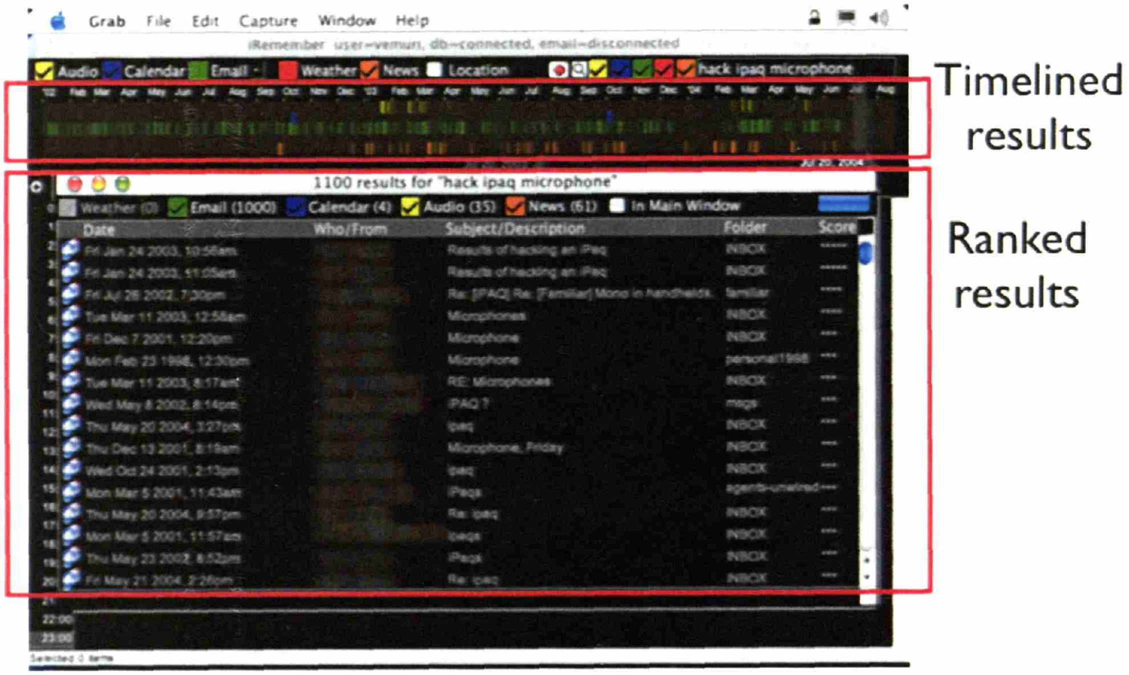
Figure 3-3: Visual interface for browsing and searching through all recordings. This view shows a multi-year timeline and a zoomed-in view of one week.

All data types are keyword-searchable. Users may enter queries into a text box seen in the upper right of Figure 3-3. Searches are performed using the open-source Lucene search engine [72]; results are relevance-ranked. Searches on news, email, and calendar entries, are bases on text extracted from each source. For audio recordings, error-laden transcripts are generated using IBM's ViaVoice® [114] speech-recognition software. Issues related to keyword-searching on speech-recognized text were presented in Section 2.3. For weather data, the textual descriptions associated with the weather data (e.g., "cloudy", "heavy snow", "light rain") are used.

Users may conduct simultaneous searches on any combination of the previously mentioned data sources (including audio); search results are presented as both a rank-ordered list and as tick marks on the timeline. Presenting results on the timeline allows users to focus on date ranges with a high density of hits or see results relative to a landmark event.

Figure 3-4 shows some example search results. The results in the figure are based on my true forgetting story told in Section 1.1. In this case, I searched through my audio, calendar, and news sources with the query "ipaq hack microphone." Results from all data types were returned; all of the high-ranking hits were email messages. The top-ranked hit was an email message from me whose subject is "Results of hacking an iPaq" and it was dated January 24, 2003. This email is a message from me to several colleagues informing them of the web page that my co-hacker and I wrote describing the iPaq hack. When I originally conducted this search, this email message helped me localize the timeframe of subsequent audio searches to late January 2003.

Clicking on a search result will automatically pan the zoomed-in view to that time. Figure 3-3 shows the week of January 24, 2003. The multiple audio recordings (yellow rectangles) on the evenings of January 22, 2003 and January 23, 2003 correspond to the audio recorded during the iPaq hacking session.


Grab File Edit Capture Window Help
iRemember user=vemuri, db=connected, email=disconnected
✓ Audio ✓ Calendar ✓ Email
✓ Weather ✓ News □ Location
hack ipaq microphone
'02 Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec '03 Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec '04 Feb Mar Apr May Jun Jul Aug
Jul 20, 2004
Timelined results
Ranked results

1100 results for "hack ipaq microphone"

□ Weather (0) ✓ Email (1000) ✓ Calendar (4) ✓ Audio (35) ✓ News (61) □ In Main Window

Date	Who/From	Subject/Description	Folder	Score
Fri Jan 24 2003, 10:56am	[redacted]	Results of hacking an iPaq	INBOX	------
Fri Jan 24 2003, 11:05am	[redacted]	Results of hacking an iPaq	INBOX	------
Fri Jul 26 2002, 7:30pm	[redacted]	Re: [iPAQ] Re: [Familiar] Mono in handhelds	familiar	----
Tue Mar 11 2003, 12:58am	[redacted]	Microphones	INBOX	----
Fri Dec 7 2001, 12:20pm	[redacted]	Microphone	INBOX	---
Mon Feb 23 1998, 12:30pm	[redacted]	Microphone	personal1998	---
Tue Mar 11 2003, 8:17am	[redacted]	RE: Microphones	INBOX	---
Wed May 8 2002, 8:14pm	[redacted]	iPAQ ?	msgs	---
Thu May 20 2004, 3:27pm	[redacted]	ipaq	INBOX	---
Thu Dec 13 2001, 8:19am	[redacted]	Microphone, Friday	INBOX	---
Wed Oct 24 2001, 2:13pm	[redacted]	ipaq	INBOX	---
Mon Mar 5 2001, 11:43am	[redacted]	iPaqs	agents-unwired	---
Thu May 20 2004, 9:57pm	[redacted]	Re: ipaq	INBOX	---
Mon Mar 5 2001, 11:57am	[redacted]	ipaqs	INBOX	---
Thu May 23 2002, 8:52pm	[redacted]	iPaqs	INBOX	---
Fri May 21 2004, 2:26pm	[redacted]	Re: ipaq	INBOX	---

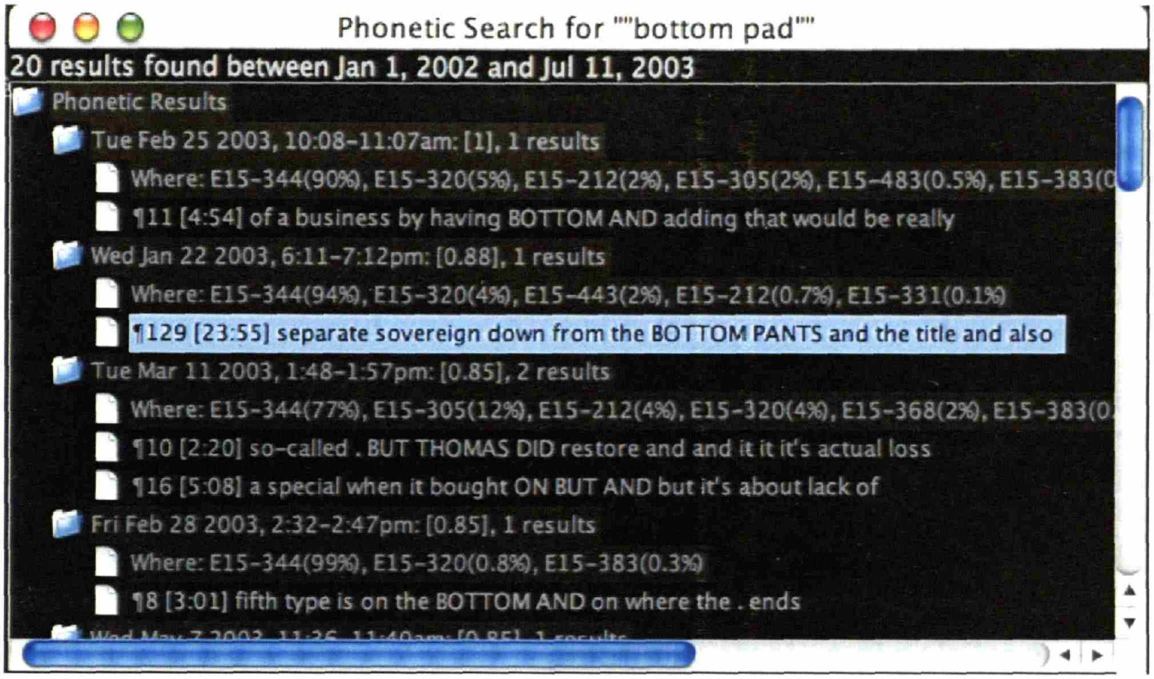
22:00
23:00
Selected 0 items

Figure 3-4: Search results are shown simultaneously as a ranked list and on the multi-year timeline. (Note: names have been blurred for publication.)

3.5.2 Collection-wide phonetic searching

When searching audio, in addition to allowing exact keyword matches (like traditional text- and spoken-document retrieval), a phonetic or "sounds-like" search feature is provided to help users identify misrecognized, out-of-vocabulary, and morphological-variant words by locating phoneme-sequence similarity. These searches can be performed across the entire collection of recordings. A screenshot of some sample results are shown in Figure 3-5. The phonetic search algorithm will be described in more detail in Section 3.7.

In the figure, the query "bottom pad" was used and the highlighted row shows a short sequence of speech-recognized words. In capital letters is the phrase "bottom pants" which corresponds to the phonetic match with the query. When a user double-clicks on the row, the short sequence (5 seconds) of audio corresponding to the transcript segment is played. This allows users to quickly skim all of the phonetic matches across all documents. One of the natural downsides of phonetic searching is it results in a large number of false positives. This facility allows users quickly explore results across all the collection before committing to a more-detailed examination of a single recording.



```
Phonetic Search for ""bottom pad""
20 results found between Jan 1, 2002 and Jul 11, 2003
Phonetic Results
    Tue Feb 25 2003, 10:08-11:07am: [1], 1 results
        Where: E15-344(90%), E15-320(5%), E15-212(2%), E15-305(2%), E15-483(0.5%), E15-383(0
        ¶11 [4:54] of a business by having BOTTOM AND adding that would be really
    Wed Jan 22 2003, 6:11-7:12pm: [0.88], 1 results
        Where: E15-344(94%), E15-320(4%), E15-443(2%), E15-212(0.7%), E15-331(0.1%)
        ¶129 [23:55] separate sovereign down from the BOTTOM PANTS and the title and also
    Tue Mar 11 2003, 1:48-1:57pm: [0.85], 2 results
        Where: E15-344(77%), E15-305(12%), E15-212(4%), E15-320(4%), E15-368(2%), E15-383(0
        ¶10 [2:20] so-called . BUT THOMAS DID restore and and it it it's actual loss
        ¶16 [5:08] a special when it bought ON BUT AND but it's about lack of
    Fri Feb 28 2003, 2:32-2:47pm: [0.85], 1 results
        Where: E15-344(99%), E15-320(0.8%), E15-383(0.3%)
        ¶8 [3:01] fifth type is on the BOTTOM AND on where the . ends
    Wed May 7 2003, 11:36-11:40am: [0.85], 1 results
```

Figure 3-5: Sample collection-wide phonetic search results for query "bottom pad"

3.5.3 Landmark searching

Similar to Ringel et al.'s interface [87], the present interface uses contextual data as "landmarks." As discussed in Section 2.1.6 landmarks can be helpful for solving certain memory problems like "forward telescoping" in which people tend to underestimate the time expired since distantly past events. Societal and environmental landmarks may be captured through sources like the news and weather. Personal landmarks may be found in email, calendar, and physical location. Other data sources are likely to have similar value; the ones mentioned here can be captured easily or passively. In addition to being able to browse such data sources temporally, they may be searched by keyword. Like audio search, results are shown simultaneously in relevance-rank order and as tick marks on the timeline. The brightness of a given mark corresponds to the sum of relevance-rank scores for all results within the time period bounded by the mark. Many results within a short period appear as dense bright areas.

Figure 3-6 shows example timeline hit-density visualizations based on searches of recent news events. The results for queries "Space Shuttle Columbia," "Iraq war," and "Janet Jackson" show that these topics appear with regularity in the news, but the news coverage intensifies at seminal moments. The "surfer shark arm" is in reference to the less-covered story of Bethany Hamilton, a surfer who lost her arm during a shark attack. Three distinct areas in the timeline, with decreasing intensity, can be seen corresponding to the first news of the attack (early November 2003), her press conference a few weeks after the attack (late November 2003), and her return to surfing a few months later (January 2004).

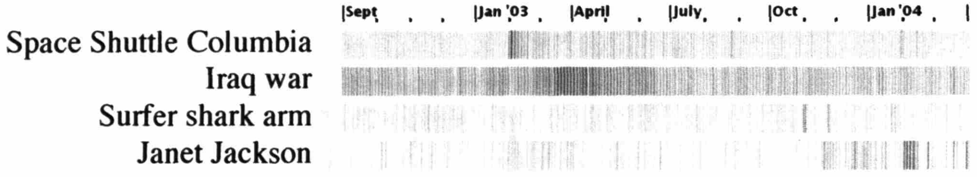


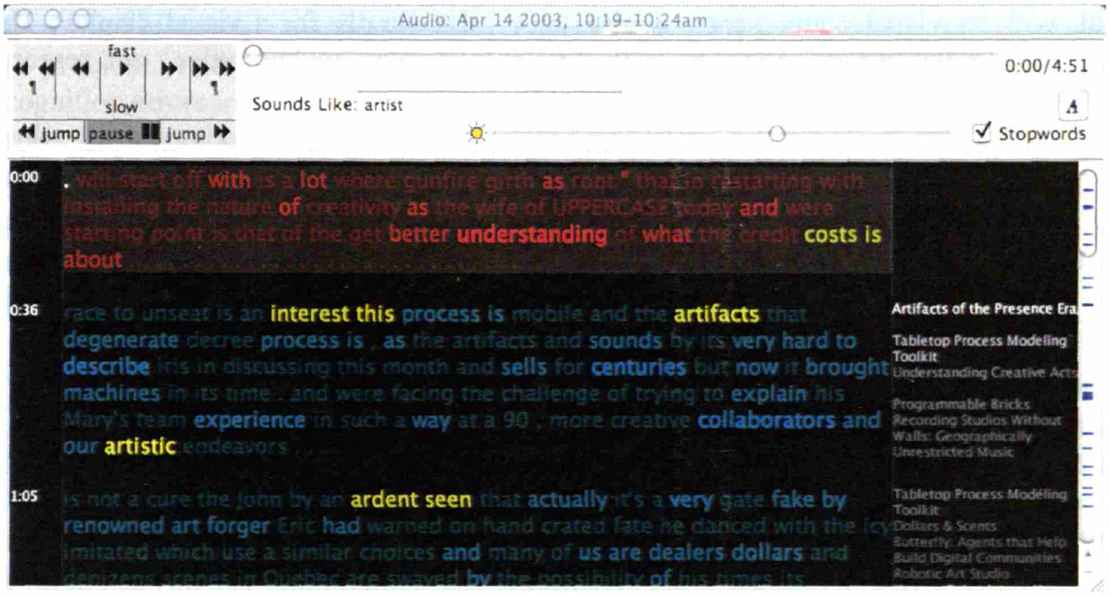
Figure 3-6: Example queries results displayed on the news timeline. Note: the visualization on the screen appears as shades of orange on a black background. These have been converted to grayscale on a white background for printing purposes.

This temporal-result-density visualization supplements conventional relevance ranking. It attempts to overcome underlying weaknesses in data and ranking algorithms by highlighting time periods possessing a mass of results. The purpose is *not* to find news, but to use landmarks in the news to help localize memories in the audio collection. It is hoped that this will prove useful in other domains where time is meaningful and search results are expected to have high recall and low precision.

3.5.4 Localizing in a single recording

Influenced by ScanMail [122], the interface described in this section addresses the problem of finding memory triggers within a single recording (Figure 3-7). Listening to long recordings is tedious and browsing error-laden transcripts is challenging [123]. Since recordings may last hours, the interface attempts to: (1) help the user find keywords in error-laden transcripts; (2) bias the user's attention towards higher quality audio; (3) help the user recall the gist of the recording; and (4) provide ways to play audio summaries that may serve as good memory triggers.

Several features are included to improve the utility of the speech-recognizer-generated transcripts. A juncture-pause-detection algorithm is used to separate text into paragraphs; this uses the speech-recognizer-reported start- and stop-times of each word. Next, similar to the Intelligent Ear [94], the transcript is rendered with each word's brightness corresponding to its speech-recognizer-reported confidence. A "brightness threshold" slider allows users to dim words whose confidence is below the threshold. This can be used to focus attention on words that are more likely to be recognized correctly. Next, an option to dim all English-language stopwords (i.e., very common words like "a", "an", "the", etc.) allows users to focus only on keywords. A rudimentary speaker-identification algorithm was included and the identifications are reflected as different-colored text (seen as red or aqua in Figure 3-7). The phonetic or "sounds-like" search feature is also provided here. Results are seen as yellow text and as dark blue marks in the scrollbar in Figure 3-7.



0:00 . will start off with is a lot where gunfire girth as root " that in restarting with installing the nature of creativity as the wife of UPPERCASE today and were starting point is that of the get better understanding of what the credit costs is about

0:36 race to unseat is an interest this process is mobile and the artifacts that degenerate decree process is , as the artifacts and sounds by its very hard to describe it's in discussing this month and sells for centuries but now it brought machines in its time . and were facing the challenge of trying to explain his Mary's team experience in such a way at a 90 , more creative collaborators and our artistic endeavors . .

Artifacts of the Presence Era
Tabletop Process Modeling Toolkit
Understanding Creative Acts
Programmable Bricks
Recording Studios Without Walls: Geographically Unrestricted Music

1:05 is not a cure the John by an ardent seen that actually it's a very gate fake by renowned art forger Eric had warned on hand crated fate he danced with the icy imitated which use a similar choices and many of us are dealers dollars and denizens scenes in Quebec are swayed by the possibility of his times its

Tabletop Process Modeling Toolkit
Dollars & Scents
Butterfly: Agents that Help Build Digital Communities
Robotic Art Studio

Figure 3-7: Interface for browsing, searching, and playing an individual recording

A sophisticated audio-playback controller (based on SpeechSkimmer [2]) capable of audio skimming is provided (upper left of Figure 3-7). Depending on where the user clicks in the controller, audio plays at normal, fast, or slow rates (pitch normalized using SOLA [50]). The control offers two forms of forward and reverse play: one skips periods of silence and the second plays only the first five seconds of each paragraph.

Finally, annotations to the transcript (seen as white text on the right of Figure 3-7) show information potentially related to the neighboring transcript text. To generate this, each section of the transcript text is used as a query to a search engine serving a corpus of roughly 800 past and present department-related project abstracts. The top-ranked project names are displayed.

Again, it was not clear which, if any, among the retrieval capabilities in these tools would prove useful for memory assistance. Among other issues, the studies that will be described in Sections 4.2 and 4.4 are designed to explore this.

3.6 Memory-retrieval tools (wearable computer)

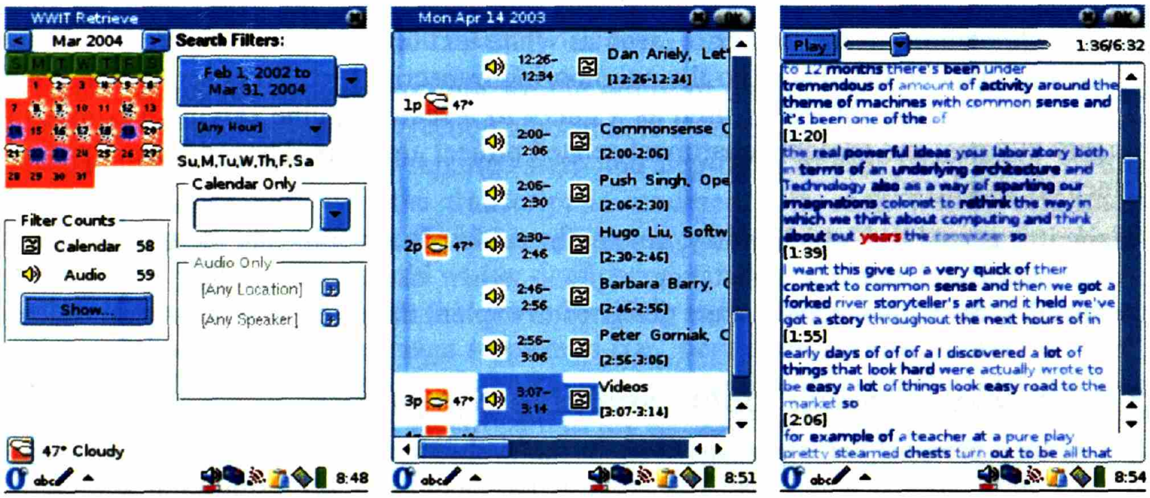
The desktop personal computer was used for the evaluations (Chapter 4) due in large part to the computational and visual demands of the searching, browsing, and skimming capabilities. But, it is also expected that users would want to perform memory-retrieval in many of the same situations they are recording. Some people regularly carry laptop computers. For these, the personal-computer-based retrieval tools would be satisfactory. The eventual goal is to move to smaller devices: mobile phones, wristwatches, key chains, jewelry, etc. One of the advantages of these is that more people are in the habit of carrying or wearing these compared to laptop computers.

The audio-recording apparatus need not be large. A microphone and either a large-capacity storage unit or a wireless network connection would suffice. But, the retrieval tool, with its visual components requires more size simply for a visual display. This assumes contemporary, commercially available display technologies, not research prototypes of miniature projection systems [14].

A subset of the key features has been implemented on the handheld iPaq PDA. These features parallel those found on the desktop computer; images of several screens are shown in Figure 3-8. The computational limits of current PDAs still require a server-based approach. In particular, real-time large-vocabulary automatic speech recognition and phonetic searching require storage space and CPU speeds just beyond the capability of current PDAs, but this is expected to change within a few years.

A shortcoming of PDAs in general is data entry. They typically do not have a mouse or other pointing device and text input via a miniature keyboards or via a numeric keypad is slow and clumsy. Speech-recognition input into such devices is becoming available, but commercially available implementations are all bound by limited vocabularies. This would be inadequate for a free-form keyword search feature.

For the purpose of a pilot study, I worked around this shortcoming. The memory-retrieval implementation on the PDA allows users to speak search queries instead of typing them. This uses the speech recognizer to translate the spoken query words into text, but the recognition is not done on the PDA. Instead a laptop computer running the ViaVoice® speech recognizer is used and the resulting text is transmitted over the computer network to the PDA in real time. This is not a long-term solution, but was adequate as part of a pilot study to evaluate the efficacy of speech-based input to a PDA for memory retrieval.



Filter Counts

Calendar	58
Audio	59

Audio Only: [Any Location] [Any Speaker]

12:26-12:34 Dan Ariely, Let
[12:26-12:34]

2:00-2:06 Commonsense C
[2:00-2:06]

2:06-2:30 Push Singh, Ope
[2:06-2:30]

2:30-2:46 Hugo Liu, Softw
[2:30-2:46]

2:46-2:56 Barbara Barry, C
[2:46-2:56]

2:56-3:06 Peter Gorniak, C
[2:56-3:06]

3:07-3:14 Videos
[3:07-3:14]

to 12 months there's been under tremendous of amount of activity around the theme of machines with common sense and it's been one of the of
[1:20]
the real powerful ideas your laboratory both in terms of an underlying architecture and Technology also as a way of sparking our imaginations colonist to rethink the way in which we think about computing and think about out years the computer so
[1:39]
I want this give up a very quick of their context to common sense and then we got a forked river storyteller's art and it held we've got a story throughout the next hours of in
[1:55]
early days of of of a I discovered a lot of things that look hard were actually wrote to be easy a lot of things look easy road to the market so
[2:06]
for example of a teacher at a pure play pretty steamed chests turn out to be all that

Figure 3-8: Memory retrieval interfaces from on the handheld computer including month-view and search interface (left), day-view (middle), and audio-playback with speech-recognition transcript (right).

3.7 Phonetic searching

In information retrieval tasks, keyword searching generally fails if there is a mismatch between the query word and the words used on the document. In text-based systems, the use of word stemmers, synonym lists, and query expansion are used to partially mitigate this problem. In speech systems, this problem is exacerbated due to the plentiful recognition errors in the transcripts; this causes frequent false-positive and false-negative results. Even so, the TREC spoken document retrieval track has ceased since recognition accuracy is the greater problem [41].

Research in speech recognition continues to make strides by reducing errors, making systems more robust to acoustic conditions, etc. Until low WERs can be achieved in the heterogeneous, noisy, multi-speaker environments typical of daily life, alternates are needed to help mitigate the error-prone nature of speech-recognizer-generated transcripts if content-based searching is to show value.

Phonetic searching is a procedure by which content-based searches are not done directly on the speech-recognizer-generated words. Instead, matches are based on words or partial words that sound like the query words. There are a variety of approaches to this and a sketch of the algorithm employed in the memory-retrieval tools described in this chapter is described below.

3.7.1 Algorithm

First, the audio recording is processed by the speech recognizer. As part of the output, the speech recognizer assigns each word a start time, a stop time, and a confidence score (the ViaVoice® "phrase score"). The recognizer outputs these as a series of words and these are separated into paragraphs and sentences as follows:

- Segment the speech into paragraphs (or "paratones") using a juncture-pause detection algorithm. This algorithm looks for consecutive words whose stop and start times are separated by at least 1400ms.
- Further break paragraphs into phrases or sentences. A silence gap of 700ms is considered sufficient to separate into sentences.

For the phonetic searching algorithm, the following steps are taken. The algorithm is a based on Wechsler et al.'s technique [119] and the relevant steps are listed below:

- Take all the words in the phrase and query and convert them to the phonetic representation by looking the word up in CMUDict [23]. If the word does not appear in CMUDict, no representation is inferred.
- For each query word, sequentially examine each phrase/sentence for sequences of phoneme similarity over a window whose width is slightly wider than the word. For example, a word with five phonemes has a window size of seven phonemes. The wide windows allows for phoneme insertions errors.
- A score is accumulated for each phoneme match between the query word and the sequence of phonemes in the sentence. If the score is above a threshold, it is considered a match. This score can also be used to rank matches.

There are two key differences between Wechsler et al.'s technique and the present one. First, Wechsler et al. use an acoustic phoneme recognizer on the audio. In the present technique, a large-vocabulary speech recognizer is used for the initial transcription an these these words are converted to phonemes using CMUDict. Second, Wechsler et al.

use a slightly lower threshold to determine a phonetic match. Their original match threshold was producing too many false positives when applied to my conversational data set. I empirically determined the higher threshold via trial and error. An example match between a query word "Singapore" and the phrase "in support" is illustrated in Figure 3-9.

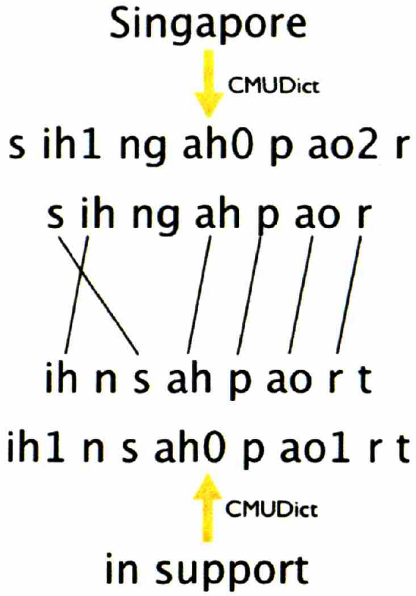


Figure 3-9: Example phonetic match between word "Singapore" and phrase "in support"

The essence of the technique is not unlike other phoneme-based or sub-word approaches [79,80]: look for phoneme sequence matches while allowing for insertion, deletion, and substitution errors. In contrast to other phoneme search systems that use a phoneme recognizer on the audio, the present technique of converts words from speech-recognition output into phonemes. The present choice to opt for this later path was pragmatic: large-vocabulary speech recognizers are commercially available and affordable; the algorithm to convert words to phonemes using CMUDict is straightforward; phoneme recognizers are not readily available; it was not a goal of the present research to build one.

No formal metric-based evaluation was conducted comparing the quality of phonetic matches against any of these other methods. Based on my informal assessment, the phonetic matches with the present technique were adequate for memory-retrieval purposes.

Chapter 4 Evaluation: Remedying memory problems

This chapter describes three studies conducted to evaluate the utility of the iRemember memory prosthesis. They are presented in chronological order. Before detailing these, I will first describe some general thoughts on the challenges of conducting memory-retrieval evaluations, give an overview of the three evaluations, and explain why these three evaluations were chosen among the many possible options.

Memory-retrieval evaluations

As with most evaluations, having controlled, objective metrics is a desirable first step. With traditional information retrieval evaluations, metrics such as precision, recall, and mean-average-precision are common. With memory studies, one simple metric is: did the subject remember a piece of information or not? The "forgetting curves" in Section 2.1.2 illustrate results of such evaluations.

With *memory-retrieval* evaluations, the metric should reflect the influence of the aid. For example, did the subject experience a measurable benefits (e.g., remember more information, make fewer errors, answer faster, etc.) with the aid compared to without? To a certain extent, Wagenaar's results [116] on using various cues (e.g., who, what, and when) provide some information regarding aided memory recall, but not for a specific memory aid or technology.

Tests on memory and memory aids in laboratory settings using investigator-crafted data sets (e.g., sequences of words, sequences of syllables, pictures, stories, etc.) can be controlled easier than tests using data collected from daily-life events. When using data from subjects' individual experiences, objectivity is not as straightforward. One person's experiences are necessarily different from another and it is unclear how to isolate variables in the face of confounds associated with the diversity of daily life. For example, suppose an investigator performed memory tests in which subjects are asked to recall details about the most recent Thanksgiving holiday. While the amount of time that has passed can be held constant across subjects, each individual's Thanksgiving will vary greatly and the vitality and lucidity of the memories might be influenced by a host of other factors: the number of times someone has reminisced about the holiday since it happened, the peculiarity of the activities compared to normal activity (e.g., traveling vs. staying at home), etc.

Two evaluation approaches to ameliorate this include within-subject designs and recruiting large numbers of subjects such that data can be categorized and normalized in some fashion. In either case, sufficient examples of authentic memory problems must be observed and remedy behavior must be documented. Doing so using subjects' real-world experiences as experimental fodder *in situ* is considerably harder compared to artificially induced memory problems in laboratory settings. Asking investigators to commit time and effort is hardly a reason to tone down an experiment, but the demands of such evaluations extend to the subjects. This remains one of the biggest challenges with long-term real-world metric-based memory studies.

In order to test the effects of a memory aid, subjects must experience real memory problems. For transience and blocking problems, this simply means someone experiencing a naturally occurring forgetting episode during their normal daily activity. As mentioned in Section 2.1.4, Eldridge et al. found that these happen regularly in the workplace [31], so memory problem frequency is not an issue. However, a big hurdle is vigilantly observing and documenting remedy behavior. When problems occur, investigators want to collect data about how subjects use the aid *in situ*. To gather this, subjects could be asked to maintain diaries, or investigators could follow subjects as they conduct their daily activity. Both methods have drawbacks: the diary method minimizes investigator burden at the cost of subject burden and can suffer from a variety of subject biases. The latter method may not have as many biases, keeps the burden on the investigator, but necessarily limits the number of subjects and situations that can be observed.

Overview of evaluations

Short of a large-scale deployment, it is unlikely that an adequately controlled evaluation over a general populace can be performed. Historically, metric-based real-world memory studies are within-subject designs on a small number of sympathetic subjects. For example, Wagenaar's [116] and Linton's [67] experiments (Section 2.1.5) involved years of vigilant personal-experience data collection that they used to evaluate their own memories.

Section 4.4 describes a similar experiment based on data I recorded with the memory prosthesis; this is the main evaluation of this chapter. However, unlike Wagenaar and Linton, I did not have my memory tested. Instead I performed memory tests on colleagues whom I recorded. The reasons for this are largely pragmatic, but it also provided an opportunity to test long-term memory of real-world experiences with non-investigator subjects (albeit a small number of subjects sympathetic to the research project).

In preparation for this evaluation, another memory-retrieval evaluation (Section 4.2) was conducted using the personal-computer-based tools (Section 3.5). For this, recordings from a set of conference talks were used instead of personal experiences This was intended to validate the memory retrieval approach, vet the memory-retrieval experimental design, and test the specific memory-retrieval implementations (Chapter 3). Evaluating with conference talks allows for a larger subject pool and is not burdened by as many confounds and challenges as evaluations based on personal experiences.

The next experiment titled "Speed listening" (Section 4.3) is not about memory per se. Instead, it validates a technique that can be used to reduce the amount of time needed to listen to recorded speech while maintaining the listeners comprehension. With the anticipated quantity of audio recorded from iRemember and similar personal data archival devices, minimizing the time needed to play audio is beneficial. The technique involves the use of pitch-normalized time-compressed audio synchronized with error-laden speech-recognizer-generated transcripts.

These three evaluations are presented in the chronological order in which they were performed. This chapter would not be complete without anecdotes of how the memory prosthesis was used by me and other volunteers. I'll start with these.

4.1 Experiences with ubiquitous recording

Given the opportunity to record anytime, anywhere, what would one record? Who would one record? Are the recordings useful to a real-world task? I have a vested interest in frequent recording and chose to record at most opportunities (within the bounds of legal, social, and human-subject protocol guidelines). Most of those recorded are within my circle of trusted co-workers and have *never* expressed hesitation about recording work-related matters. Others within the circle have been asked and have agreed to be recorded, but I felt a sense of recordee-discomfort (either by overt statements, tone of voice, or body language). When among these co-workers, I am more cautious about the circumstances and topics when asking permission to record and err on the side of not recording.

I started recording in January 2002. Most recordings took place in the MIT Media Lab; and most of these took place in or near my office or the offices of my colleagues. In total, I recorded over 360 conversations with 78 hours of audio recorded in total. The mean duration of a conversation was 12.3 minutes and the median was 6.2 minutes. Early on, most recordings were for testing purposes; I did not integrate the tool into my daily life yet. Also, at the time, I had not built the memory-retrieval software, so my motivation to record was hampered by the lack of tools to do anything with them. After I built the first memory-retrieval software, the frequency of my recordings increased. Motivation to record increased as I improved the retrieval tools. In this way, my own experiences were guiding the design in combination. Small-scale pilot studies with other users were also conducted.

Even after the retrieval tools were evolved to their current state, there were still times when I did not record. Below are some typical reasons for missing data or not recording a conversation (in rough order of decreasing frequency):

1. I turned the device on. However, unbeknownst to me until it was too late, the software crashed mid-recording.
2. I was not carrying the device at the time (recharging batteries, just walking into or out of lab).
3. I voluntarily did not record. In some situations, I felt uncomfortable asking for permission to record. Verbal consent was always required before any recording. Sometimes it felt socially awkward to ask due to the sensitive nature of the topic or because I anticipated my conversation partner might prefer not to be recorded, but would do so only begrudgingly. This happened several times per week.
4. Short or unimportant conversation expectation. Conversation partners might drop by to ask a quick question (less than one minute), perhaps about something mundane (e.g., "how late were you at lab last night?," "there's some food in the kitchen," etc.); I often did not record these. Not surprisingly, some short conversations turned out to be long; some initially unimportant conversations turned into interesting ones.
5. Absent-mindedness: I forgot to turn the device on. Sometimes I would remember partway through the conversation and retain a partial recording.

6. Other nearby conversations were taking place including people not in the protocol (and consequently did not give consent to be recorded)
7. Apathy. The life of a graduate student includes highs and lows. Fortunately, my low days were few and far-between. One some low days, I chose not to record since I did not want a record of these.

Some periods of low recording frequency align with low activity around the lab, holidays, or I was deeply engaged in solitary work. Also, only a limited number of people volunteered to be recorded, some of whom travel frequently; gaps are not surprising.

4.1.1 When I use iRemember to remember

As stated in Section 1.2.5, iRemember was not designed for a particular task or application. Through actual use, I wanted to see what tasks and situations I (and other users) would naturally chose to use memory retrieval. Though I did a lot of recording, most recordings were never used. Many of the recordings covered topics that I never felt the need to revisit. For example, a discussion of a television show.

However, there is one circumstance that I consistently feel motivated to use iRemember to find and play recordings: communications with my advisors related to an impending deadline. These typically include discussions related to a paper deadline or prior to a presentation. These conversations are often rich in content and include details that I do not want to forget. For example, when receiving comments on a paper draft, my advisors often provide a stream of comments that range from well-crafted rephrasing to insightful reorganization. The quality of the conversations is high; the content is valuable to me.

I cannot take notes fast enough and even if I could, I often cannot read my own handwriting, especially if I write hurriedly. When my advisors write notes on printed drafts, I sometimes cannot read their handwriting. If I forget something, I could ask them again, but they have busy schedules and are not always available. Also, doing so too often could make me appear inattentive. Assuming I asked again, they can also forget their original suggestions and phrasings. With most deadlines, there is too little time and too much to do. Going back to the original conversation is often the simplest path.

4.1.2 iRemember: other users

Self-reported results, especially from the primary investigator, are understandably questionable. In addition to me, two other students in the department (not directly involved in the research project) used iRemember to record select conversations. One of these students (Student Q) also used other computer-based recording tools. Student Q opts to record only a limited set of experiences: student-advisor conversations. In particular, conversations focused around an impending, time-critical task such as writing a paper: “The things that I record are things that I think ahead of time will be critical and I will need in a short amount of time. I don't record anything or everything. ...What I do make a point of recording is conversations with my advisor.” This recording strategy reflects a desire to minimize absent-mindedness problems in contrast to me: I wish to reduce transience.

Similar to Moran et al.’s observations [77], Student Q uses audio recording as a backup to hand-written note-taking: “If I write things down, it’s... like pointers into the recording and it organizes the thoughts for me." With regard to recording usage, "I often don't listen to the recording. But, to know that I have it is good. Sometimes I can't read my own writing. If I manage to write enough... that's fine. Sometimes I don't. So then I go and see what I missed, or see what I was trying to write down."

Regarding deletion strategies, Student Q—when using a non-iRemember recorder—opts not to archive recordings for space-saving reasons, but expresses no objection to keeping them. The recordings are often discarded after the task deadline, and anything important is transcribed before disposal. This point will have relevance in the forthcoming discussion about social, legal, and privacy implications (Section 5.2).

Interestingly, I have access to a much larger collection of recordings, but mainly use the recordings for the identical real-world task: retrieving student-advisor discussions in preparation for a pending writing deadline. Both of us cite common reasons for recording and usage including: limited advisor availability, the anticipated high quality of the conversations, and time-pressure of the task.

This suggests a possible application area. The student-advisor relationship is unique, but there may be analogues to the more-general supervisee-supervisor relationship. Deadlines and time-pressure are universal; quality of conversations and availability are person-specific. Though it is not clear if the utility of iRemember will extend to supervisee-supervisor relationships, this may be one area worthy of future study with memory-retrieval tools. The evaluations discussed in the remainder of the chapter were conducted in the academic setting. Future evaluations with this alternate subject pool could lead to broader applicability of the results.

4.2 Evaluation 1: Memory retrieval using conference talks

This section is a summary of an evaluation also covered in [113]. The beginning of this chapter described some of the challenges of performing memory-retrieval evaluations based on data from personal experiences. The motivation behind this evaluation is to test the memory-retrieval approach, experimental methods, and the tools on a larger and broader population without the burden of individuals recording years of conversations. An event, witnessed by many, that includes realistic, memory-demanding tasks would satisfy this. Examples of higher-participation, more-controllable, memory-demanding problem domains include students searching through archives of recorded lectures in preparation for a test, or conference attendees wishing to write trip reports. Speech from such presentation-style settings is undoubtedly different from conversations and participants are likely to be less-engaged in a such a setting; this is the tradeoff of the larger subject pool and easier data-collection benefits. While neither scenario is ideal, the memory-retrieval tools were evaluated using data from conference talks.

4.2.1 Methods

I confronted non-memory-prosthesis subjects with artificially induced memory problems based on past-witnessed events and asked them to resolve these with the personal-computer-based memory-retrieval tool. Specifically, subjects were presented with a remembering-and-finding task based on information presented in a three-day conference that occurred approximately one month prior to the test. Some subjects were speakers.

Subjects were given a questionnaire with 27 questions. Investigators designed questions with varying difficulty, some with multiple parts, and whose answers could be found and answered unambiguously. For example, a relatively simple question is "What book (title and author) does [Speaker X] cite?" and a more difficult, time-consuming example is "Which projects or people specifically cite the use of a speech recognizer?" Some questions could be answered simply by performing searches using words found in the question, some could not be answered unless the subject remembered the answer or was able to find the answer in the audio. The full list of questions is included in Appendix B.

To maximize audio quality, speakers were recorded using the auditorium's existing recording apparatus and the audio was fed into the memory prosthesis system. This approach, as opposed to using the wearable memory prosthesis device, was in anticipation of ubiquitous-computing-friendly environments that can readily support such data transfers. In total, approximately 14 hours of audio was available from 59 talks. Talk lengths ranged from two to 45 minutes.

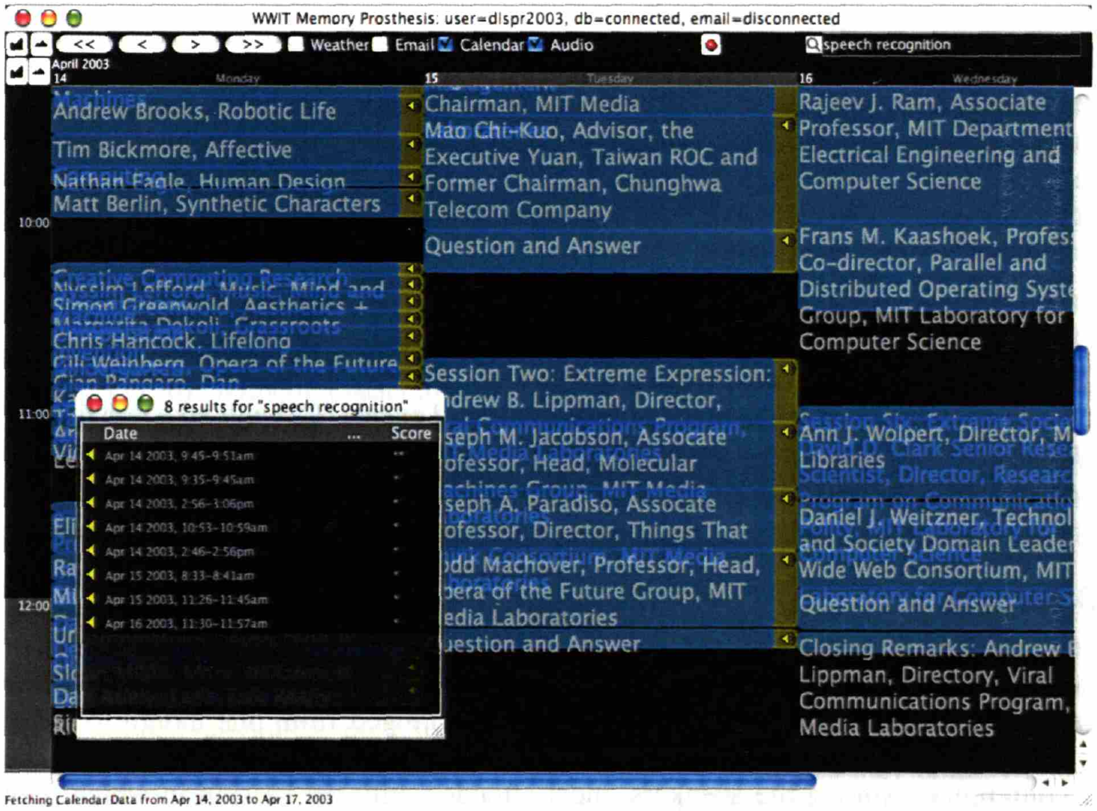
No subject attended the entire conference, but all attended at least part of it. The question dispersion was intended to ensure each subject witnessed some of the answers. It was hoped, but not guaranteed, that each subject would find and attempt questions in which they remembered witnessing the answer at the event (as opposed to knowing the answer via another source), and some memory problem occurred such that they required assistance to answer. Hence, the three experimental conditions were:

C1: Unaided: Subjects answered questions without any assistance, either by remembering the answer from the event or from another source.

C2: Aided, Witnessed: Subjects answered a question using both the memory aid and information they remembered from the event.

C3: Aided, Non-witness: Subjects did not witness the event and their answer was based on examination of data during the experiment and possibly using their previous knowledge of the department, its people, and their research in general.

Before the questions were administered, subjects were given a 5–10 minute training session with the personal-computer-based memory-retrieval software. This was an earlier version of the software described in Section 3.5. The present interface is shown in Figure 4-1. Not surprisingly, this earlier version had fewer features. Most of the missing features are not relevant to the present study; the only missing feature that might have been relevant is the collection-wide phonetic search features. The figure shows part of the agenda and recordings that took place as part of the three-day conference. For this data set, the calendar entries were copied verbatim from the conference agenda. Though not formally computed for the entire set, representative samples show a uniform WER distribution between 30–75%. Variation seemed to depend on speaker clarity, speaking rate, and accent.



Andrew Brooks, Robotic Life
Tim Bickmore, Affective
Nathan Eagle, Human Design
Matt Berlin, Synthetic Characters
Creative Computing Research
Nussim Lofford, Music, Mind and
Simon Greenwold, Aesthetics +
Margarita Dekoli, Grassroots
Chris Hancock, Lifelong
Gili Weinberg, Opera of the Future
Gian Pangaro, Dan

Chairman, MIT Media
Mao Chi-Kuo, Advisor, the Executive Yuan, Taiwan ROC and Former Chairman, Chunghwa Telecom Company
Question and Answer
Session Two: Extreme Expression:
ndrew B. Lippman, Director,
seph M. Jacobson, Assocate
ofessor, Head, Molecular
achines Group, MIT Media
seph A. Paradiso, Assocate
ofessor, Director, Things That
odd Machover, Professor, Head,
pera of the Future Group, MIT
edia Laboratories
uestion and Answer

Rajeev J. Ram, Associate Professor, MIT Department Electrical Engineering and Computer Science
Frans M. Kaashoek, Profes
Co-director, Parallel and Distributed Operating Syst
Group, MIT Laboratory for Computer Science
Session Six: Extreme Soci
Ann J. Wolpert, Director, M
Libraries
Daniel J. Weitzner, Technol
and Society Domain Leade
Wide Web Consortium, MIT
Question and Answer
Closing Remarks: Andrew B.
Lippman, Directory, Viral Communications Program, Media Laboratories

8 results for "speech recognition"

Date	...	Score
Apr 14 2003, 9:45–9:51am		**
Apr 14 2003, 9:35–9:45am		*
Apr 14 2003, 2:56–3:06pm		*
Apr 14 2003, 10:53–10:59am		*
Apr 14 2003, 2:46–2:56pm		*
Apr 15 2003, 8:33–8:41am		*
Apr 15 2003, 11:26–11:45am		*
Apr 16 2003, 11:30–11:57am		*

Fetching Calendar Data from Apr 14, 2003 to Apr 17, 2003

Figure 4-1: Visual interface for browsing and searching data in Evaluation 1.

After training, subjects were given the full questionnaire. The test was split into two phases. In Phase 1, subjects were asked to answer any questions they already knew, but without the use of any aids (Condition C1). Subjects were instructed to answer as many or as few questions as they wished and were given as much time as needed to answer these questions. In addition to collecting C1 data, this phase allowed subjects to familiarize themselves with all of the questions without any time pressure in preparation for Phase 2. A pilot study indicated that subjects did not read all the questions if under time-pressure. All subjects completed Phase 1 in less than 15 minutes.

After a subject finished unassisted question-answering, Phase 2 began in which the memory-retrieval software tool was provided. Subjects were now limited to 15 minutes to answer any remaining questions. The reasons for having a time limit included: (1) encouraging subjects to prioritize the questions they wanted to attempt; and (2) putting a limit on a subject's time commitment. Subjects were allowed to ask user-interface clarification questions without time penalty.

During both phases, subjects were asked to verbalize their thought process as they answered questions. This was audio recorded and an investigator remained in the room taking notes. Upon completion, subjects were informed about the nature of the experiment (i.e. studying search strategies), interviewed about their experience, asked to reflect on their thought process, and elaborate on any details that they might not have mentioned during the task.

4.2.2 Hypotheses

For the following hypotheses, time is computed on a per-question basis. Success rate is the number of questions answered correctly or partially-correctly divided by the number of attempts. In addition to these hypotheses, the investigators are also interested in what strategies subjects employed when remedying memory problems and what user interface features subjects found most useful.

H1a: Unaided question-answering (C1) will take less time than aided (C2 & C3).

H1b: Unaided question-answering (C1) will have a lower success rate compared to Condition C2 (aided, witnessed).

H2a: Aided question-answering of previously witnessed events (C2) will take less time than aided question-answering of non-witnessed events (C3).

H2b: Aided question-answering of previously witnessed events (C2) will have a higher success rate compared to aided question-answering of non-witnessed events (C3).

4.2.3 Results

Subjects included three women and eight men ranging in age from 18 to 50. Nine were speakers at the conference; two were non-native English speakers. Subjects' prior exposure to the research presented at the conference ranged from one month to many years and in varying capacities (e.g., students, faculty, visitors). No subject attended the entire conference. Among the speakers, most attended only the session in which they spoke. Nine subjects claimed to be fatigued during the conference and all said they were occupied with other activities (email, web browsing, chatting, daydreaming, preparing for talk, etc.) at least part of the time. Investigators classified six subjects who had prior understanding of how speech-recognition technology works.

Phase 1 (C1: Unaided Memory)

Answers were labeled as one of "correct," "partially correct," "incorrect," or "no answer." A "correct" answer satisfies all aspects of the question correctly. A "partially correct" answer has at least one aspect of the question correct but another part either incorrect, omitted, or includes erroneous extraneous information. This labeling was common for multi-part questions. An "incorrect" answer has no aspects correct. A "no answer" is one in which subjects attempted to answer, verbalizations indicated a memory problem, but no answer was submitted. Among all of these, subjects spent on average 29 seconds to read, verbalize, and answer (or choose to not-answer) a question.

Memory problems that occurred were noted. If investigators either observed or a subject verbalized a memory problem while answering, investigators classified it as one of Schacter's memory "sins" (Figure 2-3). If it was clear that a memory problem occurred, but there was ambiguity between two types of memory problems, each was assigned 0.5 points. If it was not clear if a memory problem occurred, no points were assigned. In some cases, subjects misread the question, and consequently, answered incorrectly. These were not counted. Aggregate results are summarized in Table 4-1. Investigators did not observe any of the following memory problems: absent-mindedness, bias, suggestibility, or persistence.

Answer		Problem	
Correct	47	Transience	22
Partially correct	27	Blocking	4
Incorrect	20	Misattribution	9
No answer	12		

Table 4-1: Phase 1 question-answering tallies and memory problem categorization

These results correspond to the C1 condition. The success rate (70%) is computed as the sum of correct (47) and partially correct (27) answers divided by the total number of attempts (106). Without a basis for comparison, it is difficult to say whether the question-answering performances are good or bad. Regardless, the interest with respect to the memory aid designer is the types and frequencies of memory problems. In the present case, both transience and blocking problems were found as expected, but misattribution problems were unexpectedly common. Phase 2 examines how subjects approached the task of remedying some of these.

Phase 2 (C2 and C3: Aided Memory)

As mentioned previously, the hope in Phase 2 is to observe subjects attempting to remedy memory problems (Condition C2) and understand the strategies they employ under that condition. In Phase 2, all subjects engaged in a series of searches. A search attempt begins with the subject reading the question and ends with them selecting another question (or time runs out). The memory-retrieval software recorded logs of what the subject was doing and which question they were answering. This was used in conjunction with the audio recordings of the experimental session to determine time spent on each question. Investigators further classified each question-answering attempt as either Condition C2 (subject witnessed the answer at the original event) or C3 (subject did not witness the answer). Classification was based on subject verbalizations and post-experiment interviews. Finally, attempts were classified as "successful" if the subject correctly or partially-correctly answered a new question, verified a previously answered question, or improved an answer to a previously answered question. An attempt was classified as "no answer" if the subject gave up on the search without producing an answer or time ran out. In no instances did a subject provide an incorrect answer during Phase 2. Results are detailed in Table 4-2 and summarized in Table 4-3.

A	3:22	✓ 0:56	✓ 3:40	3:30	3:30						
B	👂 7:40	👂 ✓ 5:31	1:44								
C	👂 ✓ 2:04	0:54	✓ 1:49	✓ 1:31	✓ 3:12	✓ 1:55	✓ 1:34	0:39	1:20	1:10	2:09
D	👂 ✓ 2:44	5:07	3:02	✓ 4:59							
E	👂 ✓ 1:40	1:12	✓ 3:00	1:47	3:02	4:18					
F	✓ 3:19	✓ 2:23	✓ 2:14	2:05	2:33	2:10					
G	👂 ✓ 2:39	✓ 1:46	7:11	3:51							
H	👂 ✓ 2:32	👂 ✓ 2:40	✓ 3:48	✓ 5:25	3:07						
I	1:00	✓ 4:12	👂 ✓ 1:48	1:57	✓ 1:38	✓ 2:01	✓ 2:48				
J	2:57	👂 ✓ 1:37	✓ 3:52	3:03	1:14	2:48	1:06				
K	✓ 2:01	✓ 1:21	3:22	👂 4:06	✓ 1:50	2:25					

Table 4-2: Phase 2 results. Each subject's (A–K) answering-attempts shown in sequence with time spent in subscript. 👂 = subject witnessed answer (no icon=non-witness); ✓=successful attempt (no icon=no answer). Time ran out during the last entry on each row.

	Witness (C2) 👂		Non-Witness (C3)	
	Success ✓	No Ans.	Success ✓	No Ans.
Mean	155	353	160	154
Std. Dev.	71	151	73	94
N	9	2	23	20

Table 4-3: Summary of Table 4-2 question-answering times (in seconds) and question-answering tallies (not counting C3, no answer timeouts)

The 82% success rate under C2 versus the 70% rate under C1 gives some support for Hypothesis H1b (i.e., higher success rate when aided). While also not conclusive, there is some support for Hypothesis H2b: subjects were able to answer questions more successfully in C2 (82%, 9 out of 11) compared to C3 (53%, 23 out of 43). Not surprisingly, in support of Hypothesis H1a, time spent per question under C1 was less than both C2 and C3 ($p<0.0001$). However, with no statistically significant mean differences between C2 and C3 timing, there is no support for Hypothesis H2a.

The timing data in general has caveats. Some subjects found the interface initially challenging, yet learned how to use it better over time: "I think I'm getting better at figuring out how to search the audio just in terms of thinking about things that might work." Furthermore, question difficulty was not uniform and not all subjects formulated an optimal strategy to maximize the number of answers solved. For example, some subjects intentionally chose questions out of curiosity versus optimizing their overall task-performance. Such confounding factors make it difficult to draw conclusions on timing differences. However, the timing similarity between C2 and C3 might suggest subjects have a condition-independent time threshold (roughly 4 minutes) after which they will move on whether they find an answer or not.

Observations during the experiment revealed what aspects of the interfaces (Figure 4-1 and Figure 3-7) were most valuable. Among the nine instances in which subjects were able to remedy failures, seven initiated searches by correctly identifying the talk in the calendar; the remaining two found the correct talk by keyword searching. Once in the recording, in six instances, subjects used phonetic searching to identify potentially relevant sections and limited audio playback to those. In two instances, subjects played the audio from the beginning until the answer was heard, and in one instance, the subject skipped to various points in the recording, using the transcript as a guide, until finding the answer. In one instance in which a subject was a witness but failed to find an answer, a misattribution memory problem occurred causing the subject to initially open the wrong recording. After four minutes of futile searching within the wrong audio clip, the subject gave up. In the other instance, the subject initially found the right recording, tried listening to audio from various sections (using the transcript as a guide) and phonetic searching, but to no avail.

4.2.4 Searching via speech recognition

Keyword-searching of audio collections is problematic due to the inherent errors in speech-recognizer-generated transcripts. Not surprisingly, subjects stated that poor-quality speech recognition made the task challenging.

In the present study, previous experience with speech recognition seemed to be useful. For example, one subject typed the query "brazil" to find "Breazeal" since the latter was expected to be out-of-vocabulary. Another subject focused on questions that included keywords suspected of being in the recognizer's vocabulary. Other subjects commented that an adjustment period is needed to learn the novel search features and peculiar limitations of searching speech-recognizer-generated transcripts. These subjects added that their adjustment seemed to begin within the 15-minute testing period.

The phonetic "sounds like" searching feature was used often in Phase 2. However, the out-of-vocabulary problem was still observed when subjects attempted queries with domain-specific keywords such as "BeatBugs" and "OMCSNet." The absence of these words from CMUDict [23] prevents a phonetic translation. The overall sense from subject feedback was that this was useful despite the limitation.

4.2.5 Discussion

The results give both relief and confidence that the current memory retrieval aid is a good starting point for memory retrieval of personal experiences. Subjects found answers within large collections of audio recordings, typically within a few minutes. In most cases, subjects were able to strategize ways to use the tool along with their remembrance of the past to identify the correct recording and to localize within a recording to the answer. Speech recognition, despite poor-quality transcripts, was useful for both keyword-searching audio collections and for helping subjects localize and select sections of recordings to play back.

The results, while emphasizing memory assistance, may have applicability to other audio-search domains. Subjects were able to find answers to questions from events that they did not witness, though not as accurately as in the witness-condition. Since subjects were familiar with the research in general, these results may not generalize to searching unfamiliar collections. But, there may be implications to organizational-memory applications. For example, absentees could have improved ways of quickly finding information within a recorded meeting. This example notwithstanding, memory fades over time and it is anticipated that the search process on events in the distant past will resemble that which is experienced by non-witnesses.

Despite the successful memory-retrieval examples found in this evaluation, it should be emphasized that the intention of this study is towards validating the memory-retrieval approach and refining the memory-retrieval tools. Conference talks are quite different from the events that one experiences in daily life; but it is much easier to collect data, do controlled testing, and recruit subjects with this data set compared to a data set of personal experiences. More will be said about these issues at the beginning of Section 4.4.

The next section digresses from memory retrieval and describes the evaluation of the "speed listening" technique. This helps reduce the time needed to listen to audio without sacrificing comprehension. The evaluations are presented in this sequence since they are chronological. The reader may wish to follow this ordering or skip to Section 4.4 where the evaluation of memory retrieval of personal experiences is presented.

4.3 Evaluation 2: Speed listening

The prescribed approach to memory retrieval necessarily results in large collections of recorded data, in particular, audio. Browsing audio recordings is a tedious task. If not sped-up, listening requires as much time as the original recording. For lengthy recordings, it is unlikely people would want to dedicate such time towards an information retrieval task. Observations from the conference-talk study in the previous section indicate that information-retrieval techniques (keyword searching and phonetic searching) can help subjects localize within a recording, but subjects will still dedicate considerable time to listening to the original audio. Anything that can reduce the listening time can benefit the overall time needed in the memory-retrieval search. The speed listening technique attempts to do this.

This section presents evidence in support of a technique designed to assist with information comprehension and retrieval tasks from a large collection of recorded speech. Two techniques are employed to assist users with these tasks. First, a speech recognizer creates necessarily error-laden transcripts of the recorded speech. Second, audio playback is time-compressed using the SOLAFS [50] technique. When used together, subjects are able to perform comprehension tasks with more speed and accuracy. This section summarizes previously published work [112].

To conduct the experiments, a computer program was constructed that allowed playback of time-compressed audio while visually presenting an error-laden speech-recognizer-generated transcript of the same recording. The audio was collected from a series of conference talks.

The program plays SOLAFS time-compressed audio at arbitrary speeds while displaying a transcript of that audio. The transcript appears as 18-pt. white text over a black background. While the audio plays, the program draws a line through words that have been played and highlights the current word in green. Similar to the Intelligent Ear [94], the brightness of each word in the speech-recognizer-generated transcripts is rendered proportional to its ViaVoice®-reported phrase score. Figure 4-2 shows the interface.

To test hypotheses pertaining to the subjects' comprehension of time-compressed audio with associated transcripts, five different transcript presentation styles are used:

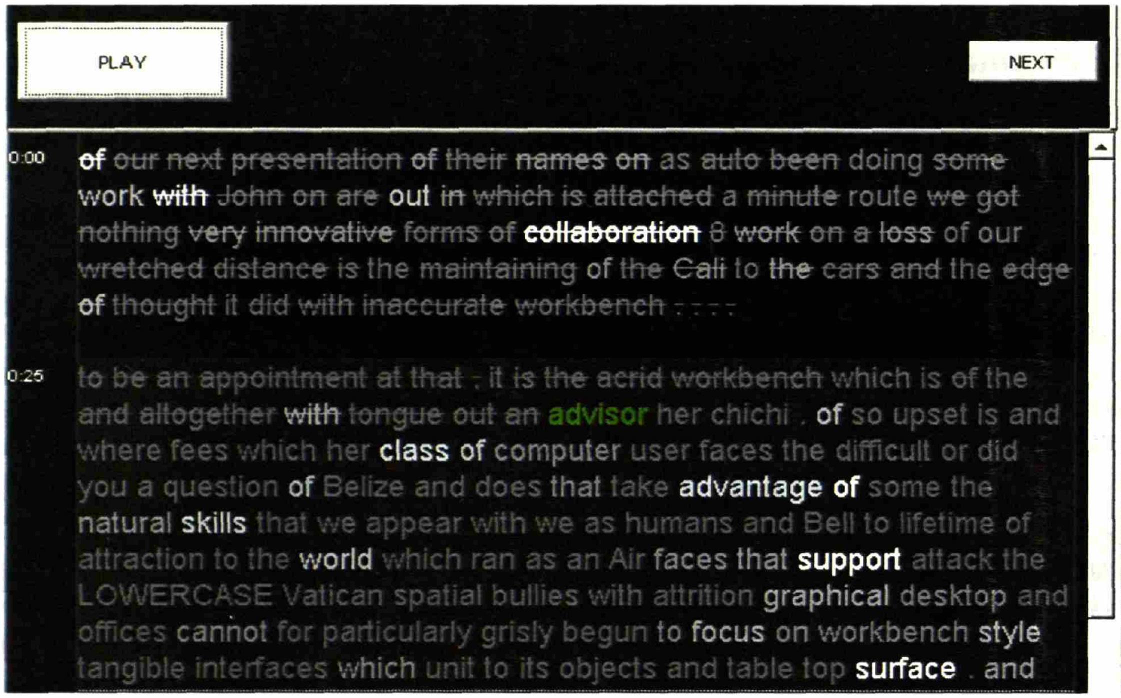
C1: Human-constructed "perfect" transcript with uniform word brightness.

C2: Speech-recognizer-generated transcript with word brightness proportional to phrase score.

C3: Speech-recognizer-generated transcript with uniform word brightness.

C4: Completely incorrect transcript with uniform word brightness.

C5: No transcript. Audio only.



0:00 ~~of our next presentation of their names on~~ as ~~auto been~~ doing ~~some~~ work ~~with John on are~~ out ~~in which is attached~~ a ~~minute~~ route ~~we got nothing very innovative~~ forms of ~~collaboration~~ 8 ~~work on a loss~~ of our ~~wretched distance~~ is the ~~maintaining~~ of ~~the Cali~~ to ~~the cars and the edge of thought~~ it ~~did with inaccurate workbench~~ - - - -

0:25 ~~to be an appointment at that~~ . it is ~~the acrid workbench~~ which is of the ~~and altogether with tongue out an~~ advisor her chichi . of so upset is and where fees which her class of computer user faces the difficult or did you a question of Belize and does that take advantage of some the natural skills that we appear with we as humans and Bell to lifetime of attraction to the world which ran as an Air faces that support attack the LOWERCASE Vatican spatial bullies with attrition graphical desktop and offices cannot for particularly grisly begun to focus on workbench style tangible interfaces which unit to its objects and table top surface . and

Figure 4-2: User interface showing brightness of individual words proportional to its phrase score.

It should be noted that style C4 transcripts are not random words. Instead, speech-recognizer-generated transcripts from sections of audio *not* corresponding to the recording are used. Next, when style C5 is presented, the program displays a string of dots whose length is proportional to the length of the audio recording and the program shows progress of audio-playback with these dots.

For the present recordings, the speech recognizer was not trained to the speakers' voices. The speech-recognizer-generated transcripts in the present data set have WERs ranging between 16% and 67% with a mean of 42% and σ = 15%. Despite the wide range and fairly uniform distribution of sample WER, it was decided not to "adjust" transcripts to a narrower band or fixed WER since it was not clear what strategy to employ to either perturb a good transcription or to correct a bad one. Furthermore, this variability seems to be an intrinsic property of large-vocabulary speech-recognition systems.

4.3.1 Hypotheses

This experiment is designed to test the effectiveness of combining speech-recognizer-generated transcripts in conjunction with pitch-normalized time-compressed speech. In particular, the following hypotheses are examined:

H1. Variation in comprehension is expected when time-compressed speech is presented in conjunction with each of the different transcript styles (C1–C5). Specifically, the transcript styles, in decreasing order of expected comprehension are C1, C2, C3, C5, and C4.

H2. The comprehension of speech played in conjunction with speech-recognizer-generated transcripts is expected to be inversely proportional to the WER of that transcript.

H3. Comprehension of SOLAFS time-compressed audio is expected to be inversely proportional to the overall speech rate expressed as words per minute (WPM).

H4. Native speakers of English are expected to be able to comprehend time-compressed audio at higher speech rates compared to non-native speakers.

The comprehension of the speech is chosen as the metric to assess these hypotheses. In the study of time-compressed audio, "'comprehension' refers to the understanding of the content of the material" [2]. Both objective and subjective measures are used to estimate this. First, a subject's subjective assessment of when they can understand a speaker under different transcript styles and time-compression rates is measured. Second, a more objective question-answering task in which subjects are tested on the contents of speech under different styles and compression-factors is performed. The next section describes this in more detail.

4.3.2 Methods

The experiment has two phases. In Phase 1, subjects are presented with three different audio samples, each taken from a single conference talk given by a single speaker. Each sample is associated with transcript style C1, C2, or C5. The order in which the samples are presented is randomized between subjects. The speech rate for all three samples averages 148 words per minute.

Subjects are presented with an interface similar to the one shown in Figure 4-2. When the subject presses the "PLAY" button, the transcript appears (or no transcript with style C5) and the audio begins playing at normal speed. The speed incrementally increases over time by increasing the SOLAFS time-compression factor. Subjects were instructed to press a "TOO FAST" button whenever they felt the playback speed was too fast to "generally understand" what was being said. This exact phrase was used so subjects would not stop simply because they missed an individual word, but would wait until the speech, in general, could not be understood. When the "TOO FAST" button is pressed, the time-compression factor is immediately reduced by 0.5 and then begins to slowly increase again. After the subject presses the button three times, playback is stopped. The software records the time-compression-factor every time the subject presses the "TOO FAST" button and averages the results.

One of the purposes of Phase 1 is to acclimate subjects to time-compressed audio in preparation for Phase 2. Previous studies suggest naïve listeners can understand pitch-normalized time-compressed audio up to a compression-factor of 2.0 and this ability improves with more exposure [82]. Subjects typically completed Phase 1 in 10–15 minutes, which is far short of the 8–10 hours prescribed by [82].

For Phase 2, subjects are presented with a series of 38 short clips of recorded speech and were tested on their understanding of those clips. To quantify subject comprehension, fill-in-the-blank style questions are asked. This provided a more objective metric compared to the self-reported comprehension assessment of the subjects in Phase 1.

The clips, when played at normal speed, have a mean duration of 20.6 seconds with $\sigma = 5.8$. Longer clips were avoided in order to minimize primacy and recency effects. As mentioned earlier, the clips were collected from a series of conference talks spanning a wide range of speakers; speakers who enunciated clearly and whose recording-quality was good were preferred. The content of the talks is mostly academic research and computer technology. The specific audio samples were selected such that there was little to no domain-specific language, jargon, and no prior knowledge was needed to understand them.

The 38 clips were presented in random order and with a random transcript style among C1 to C5. Each sample was played at a fixed time-compression factor. Audio playback speed is expressed as a time-compression factor. For example, audio played at compression 2.0 will complete in half the time of the original recording, a factor of 3.0 will complete in one-third time, etc. The first three samples were presented at factor 1.0 (i.e. original speed), the next three samples at 1.5, and in sequentially increasing factors, four samples at 1.75, 2.0, 2.25, 2.5, 2.75, 3.0, 3.25 and 3.5. Figure 4-3 shows an example distribution of the 38 sample/transcript-style pairs that might be given to a subject. Samples were presented in increasing compression-factors in order to minimize effects related to the subjects' limited exposure to time-compressed audio.

	1.0	1.5	1.75	2.0	2.25	2.5	2.75	3.0	3.25	3.5
C1	x	x		x	x	x	x		x	x
C2		x	x		x	x	x	x		x
C3		x	x	x		x	x	x	x	
C4	x		x	x	x		x	x	x	x
C5	x		x	x	x	x		x	x	x

Figure 4-3: Example of sample distribution for a single subject

Phase 2 was limited to 38 samples since pilot studies suggested subjects could complete a set of this size within the desired subject time-commitment limit. Fewer questions were assigned to the 1.0 and 1.5 compression-factors primarily due to previous results suggesting naïve listeners can understand time-compressed speech up to factor 2.0 [82].

The interface seen in Figure 4-2 was used. When the subject presses the "PLAY" button, the transcript appears (or a string of dots if transcript style C5) and the audio begins playing. When the sample finishes playing, the transcript disappears and is replaced by one to three questions about that sample. The questions ask simple, unambiguous facts about what the speaker said and do not require any interpretation or specialized knowledge. Each subject is given two practice samples and corresponding questions before the test.

Speech-rate variation among speakers suggests that time-compression factors should be normalized by a more standard speech-rate metric: words per minute (WPM). Specifically, when played at their original speeds, the audio samples in the present collection were spoken between 120 to 230 WPM with a mean of 174 and $\sigma = 29$.

4.3.3 Results

Subjects were volunteers from the M.I.T. community at large who responded to email postings on public mailing lists and to posters displayed on the M.I.T. campus. Two out of 34 subjects who participated stated they had previous exposure to time-compressed audio similar to SOLAFS. Four others said they had experience with high-speed audio, but cited examples were limited to the fast-forward feature of an analog audio-tape player, fast speech in television commercials, and some videos airing on the "MTV" television channel. Eleven subjects stated they had previous experience with speech-recognition technology, seven said they had a little experience, and 15 subjects correctly recognized the identity of at least one speaker among the recorded speakers.

Phase 1 examined subjects' self-reported maximum time-compression factor for three of the transcript styles. Figure 4-4 shows the average maximum time-compression-factor for each transcript style. Using a repeated measures, one-way ANOVA, the mean time-compression factors for all Phase 1 transcript styles were found to be different and, more precisely, C1 > C2 > C5 ($p<0.01$ for each relation). This suggests that, using the Phase 1 subjective comprehension metric, part of Hypothesis H1—which posits differences in subject comprehension among transcription styles—is supported.

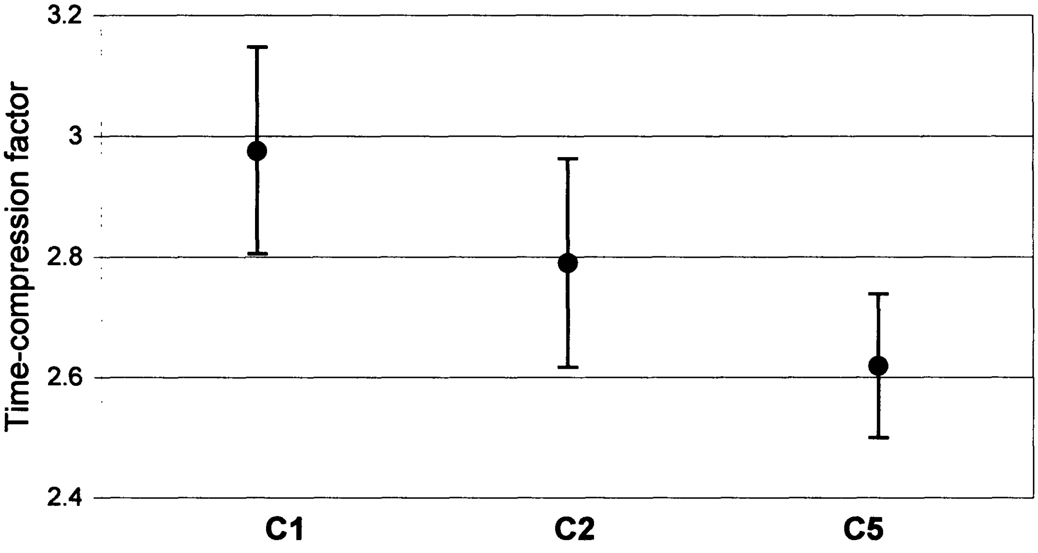


Figure 4-4: Phase 1 subject self-reported maximum time-compression factors for transcript styles C1, C2, and C5 with 95% confidence intervals.

Seven of the 34 subjects were non-native speakers of English. Across all Phase 1 transcript styles, non-native speakers averaged a maximum compression-factor of 2.47 while native speakers achieved 2.88. This difference was found to be significant ($p=0.015$) and supports the subjective aspect of native versus non-native comprehension difference (Hypothesis H4).

In the Phase 2 question-answering task, answers were judged to be correct if they indicated a subject's basic understanding of the sample's content, and incorrect otherwise. Table 4-4 shows a summary of the aggregate data for all subjects in Phase 2. The scores indicate the percentage of questions answered correctly. At compression-factors 1.0 and 1.5 each cell represents 20 to 21 data points. At all higher compression-factors, each cell represents 27 or 28 data points. These numbers do not apply to the totals, which contain the aggregate data of an entire row, column, or, in the case of the lowest-rightmost box, all 1290 data points. For samples that had more than one question, only the question subjects attempted to answer the most is included in the data. The average WPM for each cell in Table 4-4 was computed (after accounting for rate increases due to time-compression); Figure 4-5 shows subjects' question-answering accuracy for each transcript style when normalized by WPM.

Adjusting for speech-rate increases due to time-compression, the range for all samples actually played to all subjects during Phase 2 was 120 to 810 WPM. Figure 4-6 shows the fraction of all questions answered correctly across all transcript styles at each WPM decile. Significant correlation was found between WPM and subjects' question-answering accuracy (r=-0.429, p<0.0001). Hence, Hypothesis H3, which posits degradation of subject comprehension with increasing speech rate, is supported.

	C1	C2	C3	C4	C5	Total
1.0	80	85	75	90	85	83
1.5	95	90	67	81	90	84
1.75	89	79	74	81	67	78
2.0	78	78	78	74	61	73
2.25	61	81	59	59	59	64
2.5	67	59	63	46	63	60
2.75	63	55	54	37	41	50
3.0	70	54	48	15	18	41
3.25	48	22	26	11	4	22
3.5	36	26	37	7	7	23
Total	68	62	57	48	47	56

Table 4-4: Percentage of questions answered correctly at each style for each time-compression factor in Phase 2

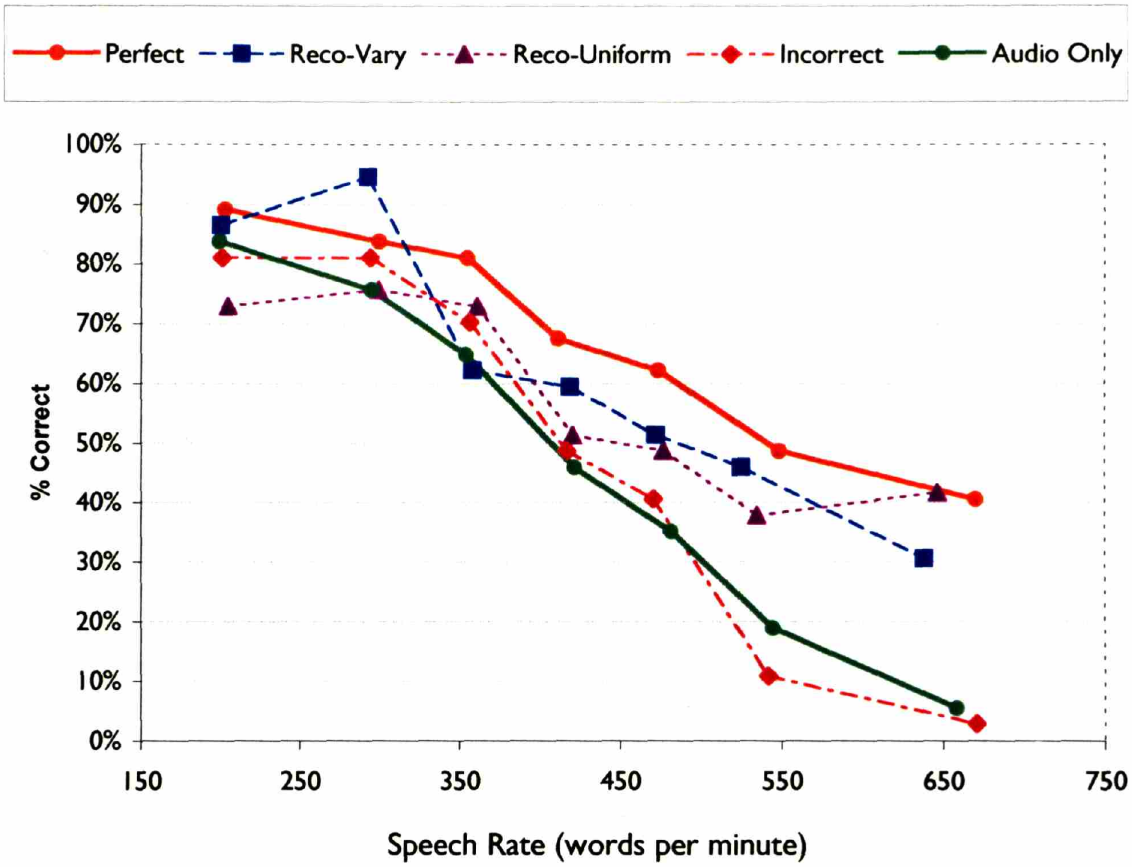


Figure 4-5: Question-answering performance vs. speech rate for the five transcript styles

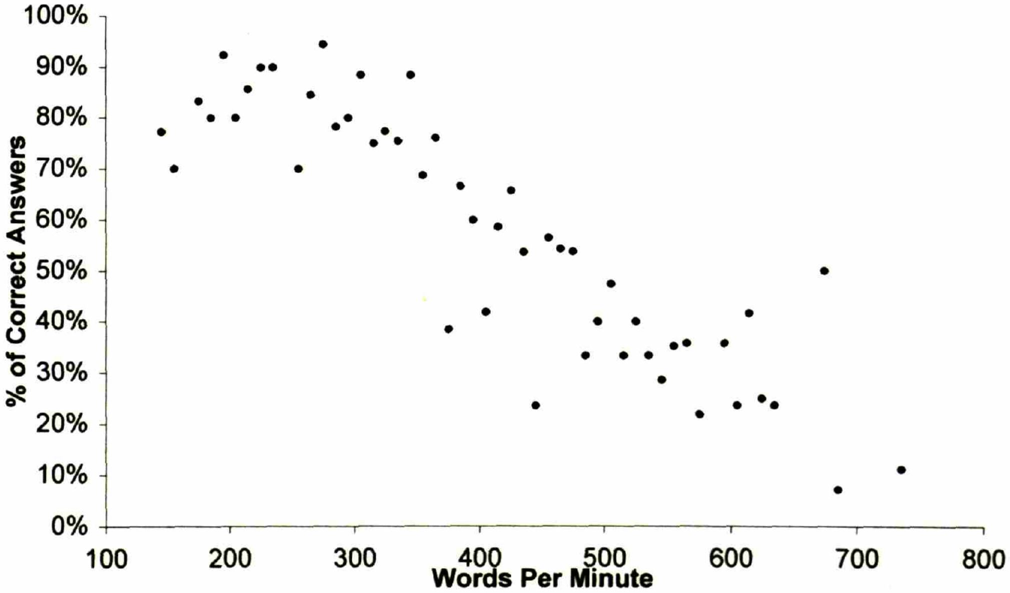


Figure 4-6: Percentage of questions answered correctly at each decile of words per minute (minimum 10 samples per decile)

Using the data from the Phase 2, question-answering task, a two-way ANOVA was conducted in which transcript style and WPM were used as independent variables and percentage of questions answered correctly was used as the dependent variable. Both transcript style and WPM showed significant variation (p<0.0001 for both), while the interaction between them did not (p=0.373). A one-way ANOVA was conducted using just transcript style as the independent variable and percentage of questions answered correctly as the dependent variable (p<0.0001). The data for this test was paired by subject. Each subject had five measures corresponding to the average number of correctly answered questions under a given transcript style. The questions for each measure were not perfectly distributed by speed and some questions were more difficult to answer than others. However, normalizing the data for speed and difficulty had a negligible effect on the overall results and such normalization has been left out of this analysis.

Table 4-5 shows the p-values obtained by comparing the comprehension scores under each transcript style with a Student-Newman-Keuls post test. Tukey and Bonferroni tests did not find a significant difference between C3/C4 and C3/C5, but otherwise yielded similar results. Figure 4-7 displays the means and 95% confidence intervals for the percentage of questions answered correctly under each of the transcripts styles.

	C1	C2	C3	C4	C5
C1	-	ns	< 0.01	< 0.001	< 0.001
C2		-	ns	< 0.001	< 0.001
C3			-	< 0.05	< 0.05
C4				-	ns
C5					-

Table 4-5: p-values associated with each pair-wise comparison between transcript styles for Phase 2 question-answering task

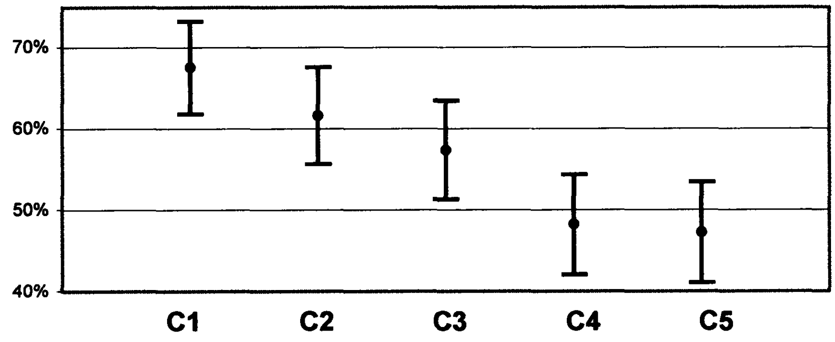


Figure 4-7: Percentage of questions answered correctly for each transcript style averaging across all time-compression factors

While these results do not support every aspect of Hypothesis H1, they do support several subcomponents. Specifically, subject comprehension of audio presented with a perfect transcript (C1) was found to be better than C3, C4, and C5 and the comprehension of C2 and C3 was found to be better than C4 and C5. No significant difference was found between completely wrong transcripts (C4) and no transcript (C5). C4 scored 0.96% higher than C5, which translates to about three questions out of 258.

In order to evaluate Hypothesis H2, correlation tests were performed comparing the WER of a given audio sample to the percentage of times subjects answered the associated question correctly across all speeds. As previously mentioned, the WER distribution across all samples was fairly uniform. Correlations were found for transcript styles C2 (r=–0.44, p=0.01) and C3 (r=–0.34, p=0.04). To ensure there were no effects related to the quality of the recordings, correlation tests were performed with styles C1 (perfect transcript, r=–0.06, p=0.73) and C5 (no transcript, r=–0.04, p=0.81). This was expected, but surprisingly, a correlation was found with the C4 transcript style (wrong transcript, r=–0.39, p=0.02). This C4 correlation implies that a transcript with no connection to the underlying audio can impact comprehension.

Finally, with respect to the differences between native and non-native English speakers (Hypothesis H4) in the Phase 2 question-answering task, Figure 4-8 shows the percentage of questions answered correctly by each group under each transcript style with p<0.005 for all native versus non-native comparisons at each style. These results suggest support for Hypothesis H4. Table 4-6 summarizes hypothesis testing for both the Phase 1 subjective tests and Phase 2 question-answering tests.

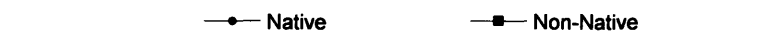


Figure 4-8: Comparison of question-answering performance for native versus non-native English speakers with 95% confidence intervals

	Phase 1: Subjective	Phase 2: Objective
H1	Supported for transcript styles C1, C2, and C5	Partially supported
H2	Not tested	Supported for C2, C3 and C4
H3	Not tested	Supported
H4	Supported	Supported

Table 4-6: Summary of hypothesis testing

4.3.4 Discussion

The perfect transcript style (C1) is tantamount to reading and, not-surprisingly, results from both Phase 1 (self-reported maximum) and Phase 2 (question-answering task) suggest this style is the best supplement to improving comprehension of speech playback. However, generating such transcripts is costly, time-consuming, and must be done manually. Using a computer speech-recognizer to generate lower-quality transcripts, like C2 and C3, can be done cheaply, quickly, and in an automated fashion.

To date, the poor transcript quality of large-vocabulary, speaker-independent recognizers has hindered more wide-scale adoption of this technology. Despite this shortcoming, the present experiment provides evidence suggesting comprehension improvements when using speech-recognizer-generated transcripts, even when laden with errors, and especially when rendered in the C2 transcript style. Specifically, comprehension of transcript style C2 was found to be better than both audio alone (C5) and a completely wrong transcript (C4). Differences between Style C2 and C3 were not found to be significant, so it is not yet clear how much Style C2's confidence-based text-brightness-rendering contributed to this, if at all.

In a worst-case scenario, a speech-recognizer may generate a completely incorrect transcript (C4). Part of Hypothesis H1 posits speech presented in conjunction with a style C4 transcript is expected to reduce comprehension compared to no transcript (C5). The supposition is that a transcript with many errors will tend to distract subjects and result in fewer correct answers. However, Phase 2 results could not support any significant difference between styles C4 and C5. Consequently, no evidence was found suggesting a completely wrong transcript would worsen comprehension compared to audio only. One possible explanation is that subjects ignored bad transcripts. Similar to the low-quality transcript abandonment results found by Stark et al. [104], some subjects in the present experiment stated that they would read a transcript for a few seconds, and elect whether or not to continue reading it based on its quality. In fact, several subjects looked away from the computer display and stated they did so to avoid the distraction of a faulty transcript.

Unexpectedly, the difference between the perfect transcript style (C1) and the brightness-coded speech-recognition style (C2) was not found to be significant in the Phase 2 objective question-answering task (though a significant difference was found in the Phase 1 subjective task). In Phase 2, a significant difference was found between C1 and the uniform-brightness speech-recognition style (C3). While it is premature to conclude that style C2 is better than C3, the evidence suggests there is some utility to visualizing text in this manner, but further investigation is needed to understand the role of brightness-coded text.

Hypothesis H1 posits comprehension variation among all transcript styles. While some aspects of this were supported (as detailed in Table 4-6), the trend suggests some of the unsupported H1 parts (specifically, C1 vs. C2 and C2 vs. C3) may achieve statistically significant variation with additional subjects.

Hypothesis H2 posits that comprehension of audio presented with a transcript will increase as the WER of the transcript decreases. This correlation was observed with the two transcript styles that had variable WER, C2 and C3. Surprisingly, comprehension of audio played with the completely wrong transcript style (C4) was correlated to the WER of the corresponding speech-recognizer-generated transcript of that audio. This non-intuitive result is not easily explained. The fact that style C1 (perfect transcript) and style C5 (no transcript) showed no correlation with WER suggest audio quality across samples was even. One possible explanation is the C4 transcript style acts as a heterogeneous progress indicator (in contrast to C5's series of dots which are homogenous). The variation might be encouraging subjects to pay attention to the audio differently. Results for this hypothesis remain inconclusive and more work is needed to understand the nature of the relationship between WER and comprehension.

Evidence for Hypothesis H3, which posits that comprehension decreases with increasing speech rate, was clearer and in agreement with Foulke's and Sticht's results [40]. Differences between native and non-native speakers (Hypothesis H4) were also found.

Collectively, these results paint an optimistic picture. Despite the fact that comprehension of time-compressed speech decreases as compression-factors increase [40], speech-recognizer-generated transcripts used in conjunction with such speech improve comprehension. In effect, the results suggest people can either save time or improve their understanding when reading error-laden speech-recognizer-generated transcripts in synchrony with time-compressed speech. The cost to provide speech-recognizer-generated transcripts is low and since very bad transcripts do not seem to confuse users, there is no apparent downside.

Though inspired by a challenge in the audio-emphasized memory-retrieval approach, the results here are not specific to memory retrieval; reduction of audio listening time can be applied to many domains. These will be discussed further in Chapter 5. This chapter continues with the chronologically next evaluation: memory-retrieval of personal experiences.

4.4 Evaluation 3: Memory retrieval of personal experiences

Two studies discussed earlier laid important groundwork for the study described in this section: Wagenaar's experiment (Section 2.1.5) detailed multi-year forgetting of salient events in daily life and the study described in Section 4.2 examined memory retrieval based on audio recorded from conference settings. The study presented in this section attempts to merge these two by examining remedy strategies based on data recorded from daily life experiences.

As stated in Section 4.2, the study of conference talks was intended to validate the memory-retrieval approach and refine the memory-retrieval tools. The eventual goal is to build a memory aid that can help people based on a much-broader set of experiences. The purpose of this study is to understand the nature of memory retrieval using recordings from personal events and experiences in daily life (as opposed to lecture data or any other approximation) and evaluate various approaches (i.e., data sources, technologies, etc.) to see what problems can be solved and how subjects try to solve them.

As mentioned earlier, one of the problems with long-term memory studies on personal data is the burdensome subject commitment prior to conducting a test. In Wagenaar's forgetting experiments, the commitment to document one to two salient events per day, every day, for six years along with sitting for periodic memory tests limited the experiment to only one subject: Wagenaar himself. He was motivated to do this; it was his research project.

Performing long-term *memory-retrieval* tests using personal data faces a similar challenge. Subjects must be willing to record their lives for years (including toting a recording device) and then have their memories tested. Who would be willing to do such a thing? Well, I am; I am motivated; it is my research project; in fact, I did. For 2.5 years, I recorded some of the normal, day-to-day conversations among willing colleagues using the iRemember recording apparatus (Section 3.3.1). But, unlike Wagenaar (who testing his own memory of past events), I tested my colleagues' memories and let them use my software as a memory aid.

There are advantages and disadvantages to testing others versus self-testing. The major advantage is that data from multiple subjects who are not the principal investigators has fewer caveats, might generalize better, and might expose interesting special cases. Some disadvantages include less data per subject and less breadth of conversation topics. Details about my data set will be covered in Section 4.4.2.1.

Very little data exists about memory problems and *repair* based on daily experiences. To a certain extent, Wagenaar's experiment provides some information about memory repair since that protocol included examples of cueing the subject with one or more of the recorded memory triggers (e.g., who, where, when, what) and asking if the subject could remember the remaining triggers. The present study explores free-form memory repair. That is, subjects are not constrained to solutions limited to a few cues. Subjects are provided with and are asked to use computer-based memory aids that include verbatim audio recordings of all asked-about events, a memory-retrieval tool allowing browsing and searching of these recordings, other personal-data (email, calendar), public data (news, weather), and any other subject-selected source that the subject wished to use to help remedy a memory problem. Based on the results from the study described in Section 4.2, the expected memory problems are transience, blocking, and misattribution.

The study of free-form memory retrieval is somewhat uncharted territory. Underlying the study is the belief that best way to explore the nature of memory retrieval and the utility of specific memory-retrieval implementations is to build it and try it.

4.4.1 Goals & measurements

The goals of this study parallel the conference-talk study in Section 4.2. At the heart of the matter: the goal is to determine if the memory-retrieval approach (i.e., use information retrieval technologies on collections of personal data) is a valid course for memory assistance. Assuming so, what aspects and technologies of specific memory-retrieval implementations (described in Chapter 3) contribute most towards success?

As stated in Chapter 1, there is little doubt that a comprehensive, verbatim archive of past experiences can help remedy memory problems. With sufficient time and motivation, the entire archive can be replayed to arrive at a definitive answer. Instead, the present study seeks to collect baseline quantitative data of various memory-retrieval tools and qualitative data on how subjects approach the task of memory remedy with such tools.

Like the conference-talk study, the basic metrics used to measure memory-retrieval performance are question-answering success and time to solution. In addition to these metrics, additional observations were made along the lines of the conference talk study. For example, does a subject's recollection of a past conversation help them find answers faster than if they did not remember?

In the conference-talk study, calendar navigation was the primary means of selecting recordings. This was attributed to the ease of navigating calendar entries spanning only a few days. Keyword searching is expected to play a larger role with the sparser, multi-year data set.

4.4.2 Methods

In a nutshell, I recorded conversations with colleagues at the Media Lab over the course of 2.5 years. After recording selected conversations in that time, I listened to all of the recordings and asked a few of the recordees questions about things we had talked about. When they could not remember the answer, they could use the iRemember software to search through our data to find the answers. I observed their retrieval efforts.

This remainder of this section describes all of this in detail including how conversational data for the memory-retrieval test was collected, how subjects were selected, how questions were constructed, and how the memory-retrieval test was administered. The reader is invited to examine these details or skip to the next section.

4.4.2.1 Conversational data collection

Collecting data from personal conversations poses many challenges. First, unlike the conference-talk study, it is not data that is usually recorded for some other purpose. More and more lecture and meeting situations are being recorded for archival purposes, but day-to-day conversations are not, yet.

The data set used for this study was the recordings between three colleagues who volunteered for the study and me. The multi-year recording efforts were described earlier in this chapter when I discussed my experiences with iRemember (Section 4.1). For the most of the recordings used in this study, the conversations had only two participants: the subject and me. All colleagues were aware of the research project, its general goals, and had given consent to be recorded and tested.

4.4.2.2 Subject selection

As stated earlier, there are advantages of testing others versus testing oneself. I could have performed self-tests as well. As the lead software engineer and designer, I am attuned to the weaknesses and subtleties in the retrieval tools. Confounds related to these (especially usability issues) could be minimized. One way I could have tested myself is similar to Wagenaar's method: Soon after each conversation, I construct questions. Months or years later, during the testing period, I have an assistant administer these questions to me and videotape the process. Unfortunately, doing so would require that I not use my multi-year collection of recordings for day-to-day memory needs. So instead, I tested others: none of whom had the retrieval software or access to the recordings.

Three colleagues were chosen to participate in memory-retrieval tests; these were selected primarily due to the number of conversations I had with them over the 2.5-year data-collection period. Other colleagues were recorded in this span, but only a few were recorded with sufficient frequency and duration to provide enough data for this study. Subjects had normal occasion to converse with me on a regular basis (i.e., outside of this study), were interested in the research, and were sympathetic to the work. In general, only colleagues who felt comfortable with me audio recording and keeping these recordings for a long time volunteered. Given their predilection towards the research, subjects were not asked about their subjective views of the technology or project.

Subjects are all researchers in the Media Lab with graduate-level academic degrees. All could be considered advanced computer users, each with over 10-years experience. Prior to the experiment, all had prior experience with speech recognition, information retrieval and all had experience both building and evaluating computer-based prototypes (hardware and software). However, aside from participating in the evaluation described in Section 4.2, none had first-hand experience searching large audio archives. All subjects are expert computer programmers using languages like C, Perl, and Java. All subjects have a deep understanding of how computer software works (including typical frailties) and are versed in Computer Science and related research fields. This group, while probably not out of the ordinary for M.I.T., is not representative of the general population.

4.4.2.3 The personal experiences data set

The audio recordings and location data were collected using the wearable recording apparatus (Section 3.3.1). Table 4-7 shows some basic statistics on the audio recordings between each of the subjects and me. In a few cases, two of the subjects and I were in the same conversation; these are counted in the tallies for both subject. No recordings included all three subjects and me.

	Number of recordings	Mean duration (minutes)	Median duration (minutes)	Total time (hours)
Subject A	58	10.2	6.2	9.8
Subject B	58	9.3	4.0	8.9
Subject C	45	11.7	7.6	8.7

Table 4-7: Basic statistics on recordings between each subject and me

The speech-recognizer-generated transcripts suffered from high WER. Some sections were better than others and variability depended primarily on the proximity of the speaker to the microphone. For most conversations, I was the speaker closest to the microphone, but I occasionally placed the device or external microphone closer to the subject or halfway between us. In the conversational-audio collection, WER seemed highest when distance between the person and the microphone was more than a few feet and if there were multiple people speaking simultaneously. The estimated WER for transcripts generated from my uninterrupted speech is 70%. Interrupted speech resulted in higher WER.

The estimated WER for the secondary speaker was very high (~100%). But, most of those errors were deletions that could be attributed to the low recording amplitude due to the distance between the secondary speaker and the microphone. So few words from the secondary speaker means the search engine has nothing to index, fewer false positives, but more false negatives. With better microphoning conditions such that the second speaker could be recorded with higher volume and clarity, one would expect the number of insertion, deletion, and substitution errors to resemble the primary speaker's.

Only one speech-recognizer voice model was used for all speakers and this model was enrolled using my voice. Under these circumstances, one would expect WER for my speech to be better than other speakers. However, this was not found to be true. Recordings from each speaker in an unrelated lecture setting were collected using a close-talking microphone (not the iPaq's); these recordings were run through the speech recognizer. The average ViaVoice® "phrase score"—which correlates to WER— for each speaker's lecture was computed. These are reported in (Table 4-8). Two of the subjects had better phrase scores than me when using my voice model. Hence, there is no evidence suggesting my speech would produce lower WER than other speakers' speech when using my voice model.

	Total Words	Phrase Score
Sunil	470	-0.35
Subject A	1261	1.48
Subject B	1214	-1.3
Subject C	736	0.85

Table 4-8: "Phrase score" for each speaker in lecture-style speech, all using the same near-field microphone and voice model

Admittedly, low-cost contemporary solutions are available that afford higher-quality microphoning. Ben Wong uses a headset noise-canceling microphone and claims this virtually eliminates secondary speakers [129]. I opted for the built-in iPaq microphone and a lavalier microphone since I did not want to eliminate audio from secondary speakers. Though the speech recognizer had very poor recognition accuracy of these speakers due to their distance from the microphone, their speech was still somewhat audible to a human listener. For memory retrieval, some audio is better than none. Future studies might benefit from all conversation participants wearing close-talking microphones. Contemporary, wireless, Bluetooth®-based microphones are one promising possibility. For the present study, adequate microphoning was the goal.

Aside from the audio recordings of conversations, subjects were asked to provide their entire calendar and email archives covering the entire 2.5-year recording period. All subjects maintained email archives over the entire span; Subjects A and B maintained calendar data for this span; Subject C's only archived data for the last year but said he would archive calendar data longer if he became a regular user of the tool.

News reports were archived automatically every night by capturing the main page of several popular news websites (CNN, New York Times, Google News, etc.); these were provided to each subject during the experiment. Weather data, including textual descriptions for the Cambridge, Massachusetts area were collected on an hourly basis throughout the 2.5-year data collection period and these were also provided.

Sometimes there were few recordings in a wide time span (e.g., one recording in a month). I may have only spoken with that colleague once that month, only recorded one conversation, or did not record the conversation for any of the previously cited reasons listed in Section 4.1. In the task, if subjects could narrow their search to that month, the task of localizing within the collection was simplified. The nearly always-on ubiquitous recording vision suggests far more data would be recorded; presumably, the higher the recording density, the harder the task. Assuming my recording behavior is typical, the volume is a reasonable approximation of the allowable recordings.

4.4.2.4 Question construction

Questionnaire construction is a time-consuming process. It takes roughly four times the duration of a recording to listen, re-listen, and extract some meaningful questions for a memory test. I listened to every conversation, identified some potentially interesting passages, and phrased them into questions. The topics encompassed a variety of discussions typical of the subjects and me. This included both research-related and personal conversations. Questions were designed such that the answers might be something useful for a research task such as writing a paper, related to day-to-day research effort, or personal interest.

Questions were designed to try to evoke transience memory problems; that is, questions that the subjects probably knew the answer at some point in the past, probably would not know the answer during the test due to memory fade, and would need assistance to answer it. In this sense, they were all hard questions. There was no way to know if the subject would experience a memory problem until the question was given. Also, there was no way to know if the subject actually encoded the relevant information to long-term memory at some point in the past. To estimate this, question topics were selected based on careful listening to the original recordings to see if there was some indication that the memory achieved some form of long-term encoding. Examples of such indications include: the subject originally spoke the answer, the subject engaged in a conversation related to the answer, the subject asked a question related to the answer, or expressed some other indication of interest in the information in the original conversation.

Questions were designed so subjects would give free-form answers. There were no true/false or multiple-choice questions. All questions had an answer that could be found among the recordings and questions were designed such that the answer was unambiguous and the entire answer could be found in a short span (roughly 30 seconds) of a single recording. In this sense, the questions were biased towards subjects' episodic memories versus their semantic memories. Subjects could ask for clarifications of the question. This happened on several occasions. A few sample questions are shown in Figure 4-9.

- What computer techniques were used to help sequence the genome?
- What does Sunil suggest is a good way to debug a deadlock situation?
- Which Star Trek episode does Sunil think is one of the best, which one is one of the worst?
- How does Singapore maintain ethnic balance in its population?
- What does MYCIN do?
- About how many words are in CMUDict? Is "Sunil" in there? Is "Vemuri"?
- According to Sunil, how many moves ahead do expert chess players plan?
- Who started the open source movement? What was the name of the project this person started?
- On some of the memory prosthesis recordings, there is a repeating tapping sound. What does Sunil think is the cause?
- What did Sunil say one should do prior to visiting the Getty Museum?
- What is Sunil's opinion on the use of common-sense reasoning to help interpret speech-recognizer-generated transcripts? On what does he base this position?
- According to Sunil, where in New York City is there an open wireless network?
- Who suggested that Sunil should build a door-closing detector? Why?
- Previously the memory prosthesis used archive.org to access past news stories. Why did Sunil choose to stop using it?

Figure 4-9: Sample questions given to subjects

Questions are biased towards things I said, not the other speaker. As stated in Section 3.3.1, the recording apparatus was positioned closer to me; the volume and quality of my recorded speech was better than that of my conversation partners. Consequently, the speech recognition for my speech was better and subjects were able to listen to my speech with greater ease than their own.

After phrasing the question, I used various combinations of words from the question to see if keyword and phonetic searching alone could be used to retrieve the answer. By design, most questions could (~90%); some could not.

4.4.2.5 Testing procedure

Subjects were presented with one question at a time and their objective was to answer this "task question" correctly. Immediately after reading the task question, they were interviewed about their remembrance of the conversation, its context, and how they would approach the task of trying to remedy the memory problem assuming it was important to do so. For the interview questions related to context of the event, subjects were asked to assess their confidence in their answer. The specific questions are shown in Figure 4-10.

1. Do you remember having this conversation?
2. What would be your answer? (Guess if you wish)
3. Would you be satisfied with this answer if it were important?
4. When did this conversation take place?
5. Where did this conversation take place?
6. Aside from me [Sunil], who else was present?
7. Is there anything else you remember about this conversation?
8. Briefly describe how you would normally go about trying to find the answer, assuming it was important, but without the software I have provided.
   - o What do you think are your chances of success this way?
   - o How quickly do you think you would find an answer?
9. Now, assume [Sunil was not available *or* you could ask Sunil]. (The phrasing of this question depended upon what the subject answered for Question 7).
   - o What do you think are your chances of success this way?
   - o How quickly do you think you would find an answer?
10. What do you think your chances of success are using the software? How quickly do you think you can find the answer?

Figure 4-10: Questionnaire given to subjects after receiving the main question, but before using software

For questions 2, 4, 5, 6, and 7, when a subject gave an answer, they were asked to asses their confidence in their answer on a scale of 0–10 with 10 meaning absolutely certain. If a subject answered Question 1 as "yes," answered the question correctly, rated their confidence in their answer high (8 or greater), and answered Question 3 as "yes," the subject was told their answer was correct and not asked to use the computer software to find the actual conversation. Under these conditions, the subject has demonstrated that there is no memory problem. Even if subjects could not answer Questions 4, 5, and 6 correctly, or gave incorrect information to Question 7, the question would still be classified as no memory problem.

After this interview, subjects were asked to use the memory-retrieval software to find the answer within the collection of recordings. Subjects were asked to speak their thoughts aloud while using the computer. The memory-retrieval software automatically logged all user interactions and the entire session was videotaped. Subjects had no time limit, but were allowed to give up at any time. Once the subject felt that they had found the answer or they wished to give up, they were interviewed again with a series of follow-up questions (Figure 4-11) similar to the earlier interview questions. At any time, subjects were allowed to give feedback and these comments were recorded.

1. Do you remember having this conversation?
2. What is your answer?Would you be satisfied with this answer if it were important?When did this conversation take place?Where did this conversation take place?Aside from me [Sunil], who else was present?Aside from what you just heard or saw, is there anything else you remember about this conversation?

Figure 4-11: Interview questions asked immediately after subject completed question-answering task

The questioning took place over the course of two weeks with each subject sitting for multiple sessions. The subject dictated the duration of each session. An individual session lasted anywhere from 10 minutes to two hours. In total, each subject spent roughly 4–5 hours attempting anywhere between 18–20 questions. Answering questions in the experimental setting can be both hilarious and fatiguing; when subjects were unsuccessful finding answers, they also found it frustrating. One subject described the task as analogous to having someone create a Trivial Pursuit®-like game, but the category is always about you.

I assessed question-answering attempts as successful or unsuccessful. A successful attempt resulted in a correct answer to the "task question." An attempt was labeled unsuccessful if either the subject did not submit an answer or submitted an incorrect guess. The success labeling was based solely on the answer to the "task question." An answer would be classified as correct even if the subject did not correctly answer any of the pre- or post-questions (i.e., when and where the event took place, who else was there, do you remembered anything else about the conversation, etc.).

I observed subjects throughout the question-answering attempt and identified memory problems. Memory problem classifications were based on observations and subject verbalizations during the interview. The possible memory categorizations included all of Schacter's seven sins and forward telescoping (for temporal assessments). Multiple memory problems were possible.

There were some unavoidable confounds. First, the process of answering one question could unintentionally improve a subject's memory and taint future question-answering attempts. As part of the normal question-answering process, subjects listed to verbatim audio from past conversations and reflected on their past. Both of these activities can strengthen memories of past events and surrounding circumstances, including events outside the bounds of the posed question. Such reflection might have the unfortunate effect of improving subjects' memories of past events as they progressed through the questionnaire. This would be reflected as improved question-answering accuracy and time-to-solution for later questions in the experiment. Essentially, for a given conversation or segment of the conversation, the memory test can only be done once. Second, subjects were expected to become more facile with the memory-retrieval tool, the nature of spoken document retrieval, and the data set as they progressed. Again, performance on later questions was expected to benefit from this.

With most experiments, a mistake might result in the elimination of one subject's data or a subset of their data. In the present case, the price can be even more costly. The data-collection effort is time consuming and subject volunteerism is low. With only a limited number of subjects and a limited number of recordings, a single mistake can result in an entire recording or recordings eliminated from all future memory tests.

4.4.2.6 Retrieval Tools

Two slight variations of the personal-computer-based memory-retrieval tools were used as part of the evaluation, which will be called UI1 and UI2 in this section. The original intent was to use only one tool throughout the study. Part way through the study, subjects provided some valuable feedback regarding features and improvements to the tool and user interface that they felt would improve their performance. Specifically, the first interface (UI1) did not allow phonetic searching across the entire collection of recordings; exact-match keyword searching was available across the collection, but phonetic searching only worked within an *individual* recording. Also, UI1 did not allow searches to be restricted to a limited date range. The usability tests on the lecture data trials (Section 4.2) did not reveal these issues; studies on general spoken-document retrieval did not give much insight regarding personal data [41,123]. The importance of these issues were only uncovered in the personal-data evaluation. This may be due to peculiarities of the personal-data situation or the higher WER among the recordings.

It was decided to modify the software to include these subject-requested features, give the modified software to the subjects for the remaining "task questions" and analyze the data separately. The second interface (UI2) is the one described in Section 3.5.

For the purpose of this task, subjects were not limited to using just the provided memory-retrieval tool. With the exception of asking me or anyone else, subjects were also allowed to use any other avenue to try to remedy the memory problem. For example, subjects could look for documents on their own computer, use their normal email or calendar client, search the web, items in their office, etc. Since the present focus is on memory remedies in the workplace setting using a computer, it was reasonable to expect that the computer might provide means outside of the memory-retrieval tool to identify landmarks.

4.4.3 Results

The results are broken into several categories. First, there are results related to memory and forgetting, but not to the software. These are garnered from the pre-questionnaires given to subjects prior to using the software. Second, there are results pertaining to memory-retrieval in general, independent of the tool. Finally, there are results related to the memory-retrieval tool and how the different interfaces impacted a subjects' ability to answer questions.

In total, subjects attempted 56 questions. In seven cases, subjects already knew the answer and did not have a memory problem. In 27 cases, subjects remembered having the conversation, but not the answer to the "task question." In 22 cases, subjects did not remember having the conversation, let alone the answer. This does not necessarily mean the memory of that conversation was not lurking somewhere in the subject's long-term memory, it just means that the question did not trigger a remembrance of the conversation. Table 4-9 summarizes the question-answering success depending upon whether the subject remembered having the conversation or not.

A "successful" memory retrieval was one that resulted in a correct answer. An "unsuccessful" retrieval was one in which the subject either gave up during the attempt or stopped their search and provided an incorrect guess. It should be noted than all "incorrect guess" cases, subjects essentially gave up on the task question and submitted a low-confidence guess.

Among the 49 task questions in which subjects had memory problems, transience memory problems were dominant with 45 cases; the remaining four were classified as misattribution problems.

	Remembered	Not remembered
Success (no software)	7 (13%)	-
Success (using software)	16 (29%)	13 (23%)
No success (using software)	11 (20%)	9 (16%)

Table 4-9: Question-answering success depending upon whether the subject remembered having the conversation

In 36 (64%) cases, subjects were successful in remembering either on their own (7 cases, 13%) or with help from the software (29 cases, 52%). It was pleasing to see that subjects succeeded even when they did not remember having the conversation (13 cases, 23%) and it was not surprising to see subjects did not succeed when they did not remember having the conversation (9 cases, 16%). Yet, it was disconcerting to observe the number of cases in which subjects remembered having the conversation, but could not find the answer (11 cases, 20%).

Figure 4-12 shows question-answering results over time. The x-axis corresponds to the amount of time that has passed between the original conversation and the memory test. The data are partitioned into two rows; the top row shows attempts in which the subject remembered having the conversation and the bottom row shows attempts in which the subjects did not remember having the conversation. Here we see that five of the seven cases in which subjects remembered the answer without using the software occurred within the six months prior to the test. We also see that most of the no-success cases correspond to conversations that were over one-year past. This includes nine of the 11 cases in which subjects remembered having the conversation, but were not able to find the answer.

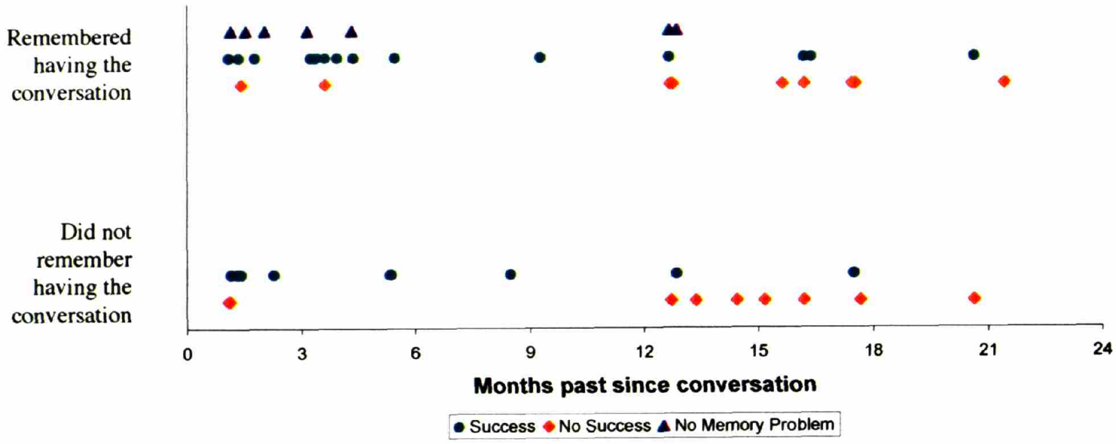


Figure 4-12: Question-answering attempts over time

Better performance on more-recent memories is not surprising, but it gives an indication that subjects' memories are involved in the retrieval process. To examine this in more detail, we look at how effectively subjects were able to predict the time bounds of a memory. For example, in the pre-questionnaire, subjects were asked when they thought the conversation took place. Subjects typically specified time ranges as a specific month, a range of months, a season of the year, a semester, etc. These were translated into a time width. If a subject said either Spring 2003 or Spring 2004, that would be a width of three months. If the conversation took place within the subject-specified bounds, that would be labeled as a "correct prediction;" otherwise incorrect. These were further partitioned into whether the subject succeeded in finding the answer to the task question. Tallies for this are shown in Table 4-10. The clustering of correct predictions corresponding to successes when the time width is two months or less gives further evidence that subjects' memories of the past events are helping with the memory-retrieval process.

Width of time prediction (months)

		<1	1	2	3	4	5	6		12+
Correct prediction	Success	5	3	4	1		1	2		1
	No success		1		1			3		3
Incorrect prediction	Success		1		2					1
	No success		1		1			1		2

Table 4-10: Subjects' prediction of time bounds. Number of questions are listed in the corresponding cell.

While time-width predictions seemed helpful, confidence in answers was not so. Again, among the 49 questions in which subjects demonstrated memory problems, subjects submitted an incorrect guess and a confidence associated with the guess for 28 of these. These are plotted in Figure 4-13. There was no correlation found between a subject's confidence in their guess and the time passed since the event (r=0.22).

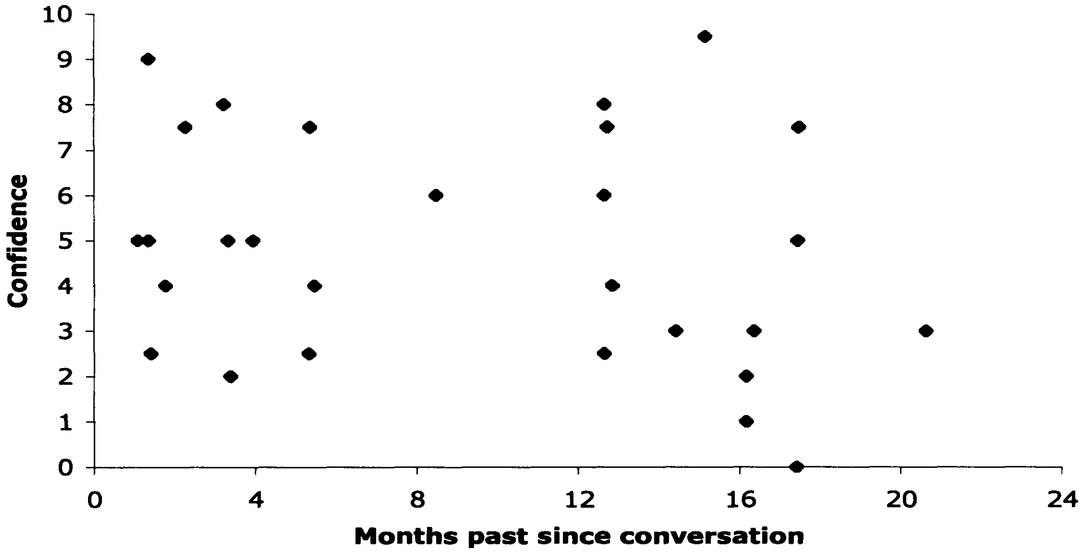


Figure 4-13: Subject reported confidence for incorrect guesses (prior to using software)

4.4.3.1 User interface differences

As mentioned earlier, the change from UI1 to UI2 was at the behest of subjects who requested certain features in anticipation of better performance. Table 4-11 shows the number of successful memory retrievals, the number of times a subject decided to give up during an attempt, and the number of times an attempt ended with an incorrect guess for both of the interfaces. UI1 had 12 successes and 12 non successes compared to UI2 (17 successes and 8 non successes). These are too few examples (especially with the confounds) to make a strong claim that UI2 is better than UI1.

	Success		Give up		Incorrect guess	
	UI1	UI2	UI1	UI2	UI1	UI2
Subject A	2	10	1	1	0	2
Subject B	4	1	7*	4	0	0
Subject C	6	6	3	1	1	0
Total	12	17	11	6	1	2

* = Subject misunderstanding of the UI1 search feature may have inflated this number

Table 4-11: Question-answering tallies for each subject when using the computer

A more-pronounced difference can be seen with the time spent per question. Table 4-12 and Table 4-13 summarize the time subjects spent per question for each respective interface. Data are partitioned based on the interface used and whether the subject was successful or if they either gave up or submitted an incorrect answer. Incorrectly answered questions and give-ups were lumped together since subjects were guessing and expressed low confidence in their answer. These results suggest that the second, and presumably better, interface (UI2) provides a small time-to-solution benefit. However, when subjects could not find the answer, they spent more time using the newer user interface. This might be because the subjects felt the newer user interface could do a better job with memory retrieval and were willing to give more effort to the task before giving up.

	Mean	Median	N
Time(success)	5:51	5:10	12
Time(no success)	10:38	8:07	12

Table 4-12: Time spent per question with UI1

	Mean	Median	N
Time(success)	5:34	4:20	17
Time(no success)	13:14	13:45	8

Table 4-13: Time spent per question with UI2

4.4.3.2 Between-subject variation

Table 4-11 also shows individual subject performance. Subjects A and C tended to be more successful than Subject B. The results suggest this was independent of user interface. Subject B did have a misunderstanding regarding the collection-wide keyword search features of UI1. I might not have made it clear to Subject B that the keyword searches in UI1 were not phonetic matches; there were exact keyword matches. Collection-wide phonetic searching was available only in UI2. Consequently, the seven "give ups" for Subject B's use of U1 might have been lower. But, Subject B's performance with UI2 (1 success, 4 non-success) suggests that something else was making the task difficult.

Based on the pre- and post-interviews, there was no evidence across subjects suggesting different remembrance of having the conversations. However, when Subjects A and C remembered having the conversation, they could cite specifics of the past conversations and the surrounding circumstances (e.g., who, where, what else). In contrast, Subject B's descriptions were more general and included references to multiple conversations with me on the asked-about topics or conversations on the topic with people outside of the study. This might suggest that Subject B's memories of these topics have become more consolidated as semantic memories whereas Subjects A and C are retained in episodic memory. The blurring of episodic details is not uncommon as memories become consolidated in semantic memory. This may further explain Subject B's lower success rate and illustrate a limitation of the memory-retrieval approach for these types of memories.

4.4.4 Discussion

Like the conference-talk study, the results here illustrate that transience memory problems can be solved via memory-retrieval. Subjects were able to achieve successes and when they did, answers could be found within a few minutes. There were times when the system could not help, but as one subject stated, "when it works, it works great." The solution presented is an early step meant to demonstrate the efficacy of the approach.

There are some important caveats of this study. The memory problems were artificially created, this study does not shed light on whether subjects would want to pay the price of vigilant recording in order to reap such benefits. This remains an open question. The timing data suggest successes can be achieved within a few minutes. Some may consider this high; some may consider it low. The reader is invited to contemplate the types of daily memory problems you experience worthy of this effort.

There were not enough conversations. In some instances, there might have been only one recorded conversation in a given month. With so few recordings, subjects could simply pick that one candidate recording. Assuming users are recording nearly continuously, the density of recordings would be much higher.

Below I discuss some of the subjective observations pertaining to search strategies and some anecdotes. A more general discussion of all of the memory-retrieval evaluations is discussed at the beginning of the next chapter.

4.4.4.1 Search strategies

The quantitative measures tell us something about subject performance using the various tools, this section discusses how subjects approached the memory-retrieval task. Some basic observations were made as subjects used the software. These can be broken down into strategies for localizing within the entire collection and strategies for localizing within a particular recording. These are enumerated in Figure 4-14.

I documented these as the subject proceeded and a video camera captured the activity on the computer screen and the subjects verbalizations throughout. For the most part, subjects employed audio search as their first choice to remedy the memory problem. In fact, despite the ability to search email, calendar, news, and weather data along with audio, all subjects turned these features off at the beginning of each session. When asked why, subjects cited the length of time to get search results from all data sources, the expectation that the result would be found only in the audio, and the difficulty in navigating a large list of results. In a few isolated circumstances, subjects did re-activate email and calendar searching.

Localizing within a collection:

1. Keyword search only
2. Collection-wide phonetic search only
3. Calendar navigation only
4. Hybrid strategy (some combination of keyword, phonetic, and calendar)
5. Using other software

Localizing within a recording

1. Listening to audio only
2. Reading transcript only
3. Listening to audio while simultaneously reading transcript of audio
4. Listening to audio while simultaneously reading transcript of another section
5. Skimming (using speech skimmer)
6. Skimming (based on speaker ID)
7. Skimming (forward, pseudo-random)
8. Skimming (all over, pseudo-random)
9. Removing stopwords
10. Changing brightness
11. Using ranked phonetic matches
12. Using tick-marks in scrollbar
13. Scanning over phonetic search hits (yellow text)

Figure 4-14: Basic activities observed during subject use of software

Below is a list of some other general observations from the present study related to search strategies:

- Within-recording localization strategies were similar to those in the conference-talk study (Section 4.2). Considering the audio "document" visualization interface (Figure 3-7) was mostly unchanged between the evaluations, this was not surprising.
- Keyword search was the preferred mechanism for collection-wide audio search. This was contrary the conference-talk study where calendar-navigation was the primary choice. Moreover, with no temporal or landmark cue to use, keyword search is often the only remaining choice.
- Landmarks helped with time-localization when used, but subjects employed simple mechanisms like calendar, email, web search to find landmarks. The more-sophisticated features in the temporal-density landmark search feature (Section 3.5.3) were not needed.
- Accurate speech recognition makes the task easier.
- Less-vivid remembrance made it harder. This was based on subject's answers in the pre-questionnaire indicating vivid details of the conversation, even if they did not remember the answer to the specific question.
- Misattribution, both in answering the "task" question and in the pre- and post-questionnaires, led subjects astray. For example, looking in the wrong time period.

- Subjects stated that multi-tasking was a helpful way to search through the data. For example, a subject would let an audio recording play while simultaneously reading another transcript or initiating another search.

In any search task (memory or otherwise), a wrong turn leads to penalties. The time results give some objective evidence towards this nature of the penalty for memory retrieval. When searching collections, subjects would formulate an initial search strategy (e.g., keywords within a recording, looking for a landmark in either email or calendar entries, etc.), and try it for some time. Initial failure might lead to minor variations on the query. For example, choosing slightly different keywords, variations on the original keywords, or different Boolean operators. If that didn't work, they might try another path or just give up. I was observing in real time and I could often predict if the subject would succeed or fail at a question within about one minute. This was not tracked formally as I only realized this after many examples had passed. When I informed subjects of this, they requested that I include a feature that would tell them quickly if they were on a doomed path. How to do this remains an open question.

4.4.4.2 Anecdotes

In addition to observing trends, several specific cases were instructive and entertaining. These are listed below in no particular order, but simply to document these cases.

- When keyword searching of the main topic fails, subjects searched for what they thought was another topic in the same conversation. One subject remembered having the conversation, but was not succeeding in localizing within the collection using keyword search. The subject attributed this to poor speech recognition. Instead of giving up, the subject remembered another topic in that same conversation and started to search for keywords related to this second topic. The subject found the audio associated with this secondary topic and then skimmed the audio to locate the answer to the original task question.
- Three times, subjects remembered the conversation took place soon after a scheduled seminar or thesis defense. The subjects searched their email (using both iRemember and their own email clients) for the talk announcement. Once found, they used that as a landmark and focused their attention on only the audio recordings soon after the talk.
- In one case, a subject remembered a conversation took place soon after I had returned from a conference. The subject used a web search engine to find the dates of a conference and focused on recordings occurring soon thereafter. Similarly, one subject remembered a conversation took place soon after his vacation and used that as a landmark.
- After participating in the study, one subject felt inclined to pay closer attention to his conversation partners; he felt his memory of some of the important aspects of the conversations was disappointing. Participating in the study revealed weaknesses in his memory that he was previously unaware.
- Some subject frustration in the study could be attributed to shortcomings in the user interfaces. The interfaces are research prototypes and admittedly had some usability

shortcomings (e.g., speed, design, etc.). However, one subject also express frustration because of inability to produce the answer unaided "I should know this and I'm disappointed that I do not."

- In the single funniest question among them all, one of the subjects, expressing skepticism about the memory-retrieval approach, had stated in a recorded conversation nearly two-years past that he felt it was "highly unlikely" that we would be able find this conversation years later; some colorful phrasing was used. Feeling up to the challenge, I phrased this into a question in that subject's test: "In the conversation where [Subject] and Sunil were talking about "[colorful phrase]," what did [Subject] say it was unlikely Sunil could do?" The subject found the answer.

Chapter 5 Design implications

In this chapter, lessons learned from the evaluations described in the previous chapter are synthesized. The first section revisits the evaluations from the previous chapter. The second covers the social, legal, and privacy implications of the technologies introduced herein. Finally, the last section introduces a smattering of potential future work areas related to memory retrieval and general computational memory assistance.

5.1 Implications from the evaluations

This section revisits the evaluations and expands on some issues brought up earlier. Localizing is the key to a successful search task. As anticipated, subject's prior experience with the data and their recollection of how a conversation/talk proceeded and/or their expectation of where certain data is likely to appear helps in the overall information retrieval task. This section drills into a variety of issues raised from the memory-retrieval evaluations.

5.1.1 Spoken document retrieval as part of memory retrieval

Keyword searches of audio was one of the common search strategies. It played a bigger role in the personal-experiences study, but was still evident in the conference-talk study. Thanks in large part to the WWW and web-based search engines, most computer-savvy people have become familiar with text-based information retrieval. Searching audio is somewhat novel and most subjects felt that some adjustment time was needed to better understand how to best use the tool.

In the evaluations, audio search was not the only means, but there were cases in which subject could neither recall the conversation nor anything about its context (who, when, where, etc). When this happened, the *only* options are keyword search or systematically listening to the entire archive. The latter remains undesirable. Even if there were other search paths (e.g., finding landmarks, browsing to a date range, etc.), subjects would still try searching audio; the cost is not high to do so: type a few words, await some results.

With audio as a substrate, memory retrieval is inextricably tied to spoken document retrieval, including all of its benefits and pitfalls. Section 2.3.1 elaborates on some of these. Given the recent progress in speech recognition technology, it is fair to assume that it will continue to improve and become more robust under the harsh conditions of everyday life. Along with such improvements, one would expect that keyword searching would increasingly supplant other forms of navigation and search. With regard to localizing within a document, one would expect similar benefits to phonetic searching as WER decreases.

Until higher accuracy is achieved, the visualization, phonetic search and other techniques necessarily take on important roles in the retrieval process. One can think of these as crutches that mitigate the shortcomings of an intrinsically error-prone technology. The "speed listening" experiment (Section 4.3) illustrated how error-prone transcripts can be better than audio alone, and having the transcript facilitates other forms of user cleverness. For example, the subject who type the query "Brazil" when searching for the word "Breazeal."

Subjects can also demonstrate clever approaches based on their remembrance. In general spoken document retrieval tasks, searching for topic-specific keywords is a sensible choice. But, the keywords might not have been spoken in the original conversation. For example, I can have a conversation about baseball without ever saying the word "baseball." It is also possible that the keywords were spoken, but misrecognized. In these cases, keyword search is prone to failure. One's memory can help. If subjects remember other parts of the conversation or its context, they can use this information to formulate alternate search queries. An example of a subject doing just this was cited in the evaluation of personal experiences (Section 4.4).

Next, even a well-thought-out query might not get good results, or can lead to time-wasting paths. Even with text-based search on large corpora, one can spend a great deal of time looking for an answer to a question and not find the answer until the query is phrased in a way that will get the results. To digress for a moment, not long ago a colleague was having difficulty modifying a Fortran program so that it would communicate with his C program via "command line arguments." After several minutes of web searching queries like "Fortran command line arguments" and variations thereof, he called upon me for advice. When I suggested the query "Fortran argc argv," my colleague seemed puzzled, but obliged. The top web search result gave the desired answer. Now even more baffled, my colleague asked me what inspired me to select obscure words like "argc" and "argv." C programmers often use the variable names "argc" and "argv" to designate command line arguments in their code. Knowing this helped me formulate a query that resulted in more useful results.

Similarly, with audio, formulating the right query can mean the difference between success and failure. Again, citing the "Breazeal" for "Brazil" example, we can see how this understanding of the technology and domain can make a difference. Even so, I have done hundreds of memory-retrieval searches using iRemember and I still do not fully grasp the nuances of constructing better queries. The learning curve for this might be steeper than text-based search.

5.1.2 Audio Summarization

Unlike most text documents, audio recordings do not have titles or other succinct summaries of their content. With the present tools, the date of a recording as well as any associated context (calendar, email, location, weather, etc.) is provided, but these can only offer limited clues towards the content. Providing users some form of summary that can give stronger clues towards the contents, especially one tuned towards the query, is likely to go a long way towards helping them refine their search paths.

In textual searching, this is often done by either a computer-generated abstract of the document or extracting excerpts focused on regions of keyword matches. It is not clear how a conversational-audio abstract can be created. A conversation may cover many topics and include digressions. Synthetic News Radio found ways to segment speech into topics, but this was limited to the broadcast news domain and depended upon text corpora of current news topics [33]. Segmentation suffers greatly when using speech-recognition transcripts alone. The higher error rate of spontaneous, conversational speech and lack of an independent textual analog to the conversations suggests this technique could not be applied.

However, topic extraction can be done with audio. For example, similar to text-extraction, the regions of audio with the highest density of keyword matches can be stitched together to produce a short audio clip. This could be provided along with the transcript words as part of the search results list. This feature was considered, but the required software-engineering effort was beyond the scope of the evaluations.

5.1.3 Topic identification

A prose summary of a recording, while helpful, may not be the only useful, succinct description. Having some general sense of the topics discussed in a recording or even in regions is useful. The high WER of the transcripts makes this difficult, and the design criteria for the single recording visualization interface (Figure 3-7) aims to help users make these deductions. But, what if the computer can make these deductions automatically?

Though natural language parsing is not likely to succeed due to the high WER, examination of the words (including the errors) might be fruitful. Eagle et al. [29] attempted this using common-sense reasoning techniques and found some successes. Based on this, one of the co-authors, Push Singh, and I tried the same technique with recordings from my data set. Sadly, we were not as successful. The knowledge contained in the OpenMind common-sense database is general knowledge analogous to semantic memory (i.e., world knowledge, concepts, etc.). Most of my recordings were more-specialized topics (e.g., my research) that are unlikely to appear in a repository like OpenMind.

5.1.4 Usability issues

Subjects expressed dissatisfaction in both the interface and the performance (i.e., time it takes to perform certain operations). Though subjects were instructed that the software was an experimental prototype and the test was not a usability study, usability issues seemed to influence subjects' choices of search paths. For example, email search was computationally demanding and subjects would disable searching that source so results from the other source would be returned faster. While these are legitimate concerns for both researchers and practitioners, and some effort was given to usability and performance issues, perfection of aesthetics, usability, and performance were not explicit goals of the work. Part of the reason is the novelty of the memory-retrieval task and minimal examples to model after. A more-rigorous usability study prior to the main study would likely have helped with interface issues. To a certain extent, this was the purpose of the pilot studies and conference-talk evaluation (Section 4.2). A best effort was given, but the size and complexity of the software required prioritization. The experimental protocol was designed around subjects familiar with limitations of research prototypes vs. products. Nevertheless, the imperfections did seem to play a more important role that I originally expected.

5.1.5 Landmarks

Section 2.1.6 discussed the value of landmarks, Section 3.5.3 described iRemember's implementation. In review, the retrieval tool offered a variety of ways to search for landmark events including email and calendar for personal landmarks, and news and weather for public landmarks. This was put to evaluation as part of the personal-experiences study (Section 4.4). To my disappointment, none were used with regularity. Subjects tended to prefer to see the entire list and filter on their own.

This is not to say that subjects did not use landmarks, they just chose not to use timeline-density rendering of search results. The preferred mechanisms were calendar and email search to find the relevant entry. For example, if the subject remembered the conversation took place soon after a seminar, they would search for the seminar announcement in email. Once the time bounds were established, the subject would either select recordings around that time in the calendar view or do a keyword search and only select recordings within their expected time bounds. In the end, this meant that the solution provided was over-engineered to the task. Something simpler would have sufficed.

5.1.6 Fruitless search paths

Designers of search tools dread, but must reckon with helping users avoid fruitless search paths. These are trajectories in an information search path that users decide to follow but will not result in any useful information. Users are tempted down such paths because they feel an answer is forthcoming. In some cases such paths are a result of an ineffective query. In other circumstances, the information may not exist in the database despite the user's expectation of its presence. For example, a user searching a low-tier encyclopedia for the capital of a little-known country (e.g., Tuvalu[1]) might not find it because the particular encyclopedia vendor did not include the country. Fruitless search paths cause time wastage.

One's memory can serve cause interference and even instigate the pursuit of fruitless paths. For example, if a person remembers a conversation, searches for it, but forgets that it was never recorded. Conversely, a person can search for a recording, fail to find it, and assume it was not recorded. Both situations are undesirable.

In the personal-experiences study, there was one notable instance of a subject remembering a particular conversation where he expected to find the answer to the posed question, but the conversation was not recorded. This sent the subject down a fruitless search path. The subject verbalizations indicated a strong belief that the answer was in a conversation within a particular time frame, but since the conversation was never recorded, the subject spent several minutes sequentially going through other recordings within that time bounds to no avail. Eventually, the subject did find the answer in a different conversation. A similar instance was found in the conference-talk study, but the subject gave up.

This introduces a shortcoming of the memory aid in general: users may remember having the conversation, but may forget if the conversation was recorded. Having a more-

[1] According to the CIA World Factbook, Tuvalu is a 26 sq. km. Island in the south pacific whose population is roughly 11,500 and whose capital is Funafuti

comprehensive archive would help remedy this problem at the cost of including more superfluous recordings.

5.1.7 Misattribution

Both the conference-talk and personal-experiences studies were designed to evaluate transience and blocking memory problems. Instances of blocking were found in the conference-talk study, but none were observed in the personal-experiences study. However, misattribution was found in both. In the personal-experiences study, this was mostly manifested as incorrect answers to the pre- and post-questionnaire. Subjects would associated a memory with the wrong time, the wrong place, the wrong person, or would include incorrect information when asked if they remember anything else about the conversation. In some cases, subjects discovered the misattributions through the course of answering the question using the recordings. For example, one subject was convinced a particular conversation happened on a given day and described detailed circumstances of the conversation and how he remembered it; he was perplexed to discover the conversation took place one month earlier than expected.

Misattribution problems can lead subjects to fruitless search paths and the misattribution need only correspond to one aspect of the memory (who, where, when, etc.). Subjects' high confidence in the erroneous memory is a recipe for further time wastage. It is not clear how the memory prosthesis can help remedy these memory problems or how it can help people avoid these pitfalls. Misattribution is characterized by a person erroneously thinking they remember correctly. This remembrance would need to be broken down first and it is not clear how a computer search tool would recognize this.

5.1.8 Localization within a document vs. within a collection

Whittaker et al. suggest the bigger challenge in spoken-document retrieval is localization within an audio "document" [123] as opposed to within a collection. This held true in the conference-talk study where subjects were able to quickly identify the correct recording in most memory-retrieval cases. But it did not hold in the personal-experiences study. In the latter, subject observations indicated that the greater problem was localization within a collection.

First, for the reasons discussed earlier in this chapter, finding the right recording was not always straightforward. Whittaker et al.'s stance might be rooted in the study of broadcast news, which typically has a lower WER compared to spontaneous speech. With a lower WER, there is a higher chance of keyword match when doing collection-wide searches. Also, the spoken document community claims success at the problem [41]; TREC SDR no longer takes place. TREC SDR also perform evaluations on broadcast news and their stance is based on evidence suggesting that improvements to SDR are now mostly dependent upon improvements to speech recognition and not information retrieval techniques.

Another possible explanation for the higher burden of collection-wide localization is the penalty incurred once subjects select a single recording for deeper investigation. In the personal-experiences study, once a subject committed to a particular recording, they invested time in that recording. Subjects would easily spend several minutes on these deeper per-recording investigations. This included playing and replaying multiple segments within the entire recording as part of the search task. The particular implementation exacerbated this a bit in that it incurred a time burden whenever a new audio document was opened; it could take tens of seconds to download and render the interface (Figure 3-7). The next section discusses how the collection-wide phonetic search feature emerged and how it tries to address some of these localization challenges.

5.1.9 Collection-wide phonetic searching

As mentioned in Section 4.4.3.1, the personal-experiences study started with a single interface for the memory-retrieval tool, but a second was created in response to subject feedback. The major difference between the two interfaces (UI1 and UI2) was the addition of collection-wide phonetic search in the second interface (UI2). This feature was meant to partially address the localization challenges described in the previous section.

Phonetic searching was available to subjects already, but it was only included as part of the single document playback and search interface (Figure 3-7). The reason for this limitation is the CPU processing demands of the phonetic search algorithm. On my full data set that includes over 70 hours of audio, a simple phonetic search takes roughly 30 seconds on a reasonably fast desktop computer (dual 1.42 Ghz, PowerPC G4 Macintosh). On the smaller data set of each subject (roughly 9 hours), results were computed in just a few seconds, even on a slower machine.

Subjects felt the looser match criterion of phonetic searching (compared to exact keyword match) was more attuned to the way they wanted to search the high-WER transcripts. Also, when examining phonetic search results within a recording, subjects would often just play short segments of audio surrounding a phonetic search match to determine if it was a valid section. When subjects heard something in the speech or saw something on a transcript related to their query topic, they were adept at identifying that.

The collection-wide phonetic search tool (Section 3.5.2) encompassed all of this. The interface that listed all of the audio segments from all documents simultaneously allowed for simpler visual skimming of these results. The interface also allowed users to play the short audio segments without committing to opening an entire recording. On the whole, subjects were positive about the feature and the timing data from the personal-experiences study indicate that subjects were willing to dedicate more time to the task with this facility. But the limited number of questions and intrinsic confounds of personal-experiences memory-retrieval tests prevent claims on improved question-answering success.

5.1.10 Normalizing personal experiences

As stated at the beginning of Chapter 4, confounds are intrinsic in memory-retrieval studies. It is difficult to control for the varying experiences people have, the varying remembrance they have of shared experiences, etc. One of the confounds in both the conference-talk and personal-experiences studies was the varying difficulty of questions. A possible way to put a degree of control on this is by assigning a difficulty score per question and using this to normalize results. I present a proposal metric here. This can be broken down into subject-independent and subject-specific factors.

A subject-independent difficulty metric can be assigned to a question based on the number of paths to the answer the question constructor can identify using keyword and phonetic searching. Here, a path is a sequence of steps using the memory-retrieval tool (e.g., keyword search, followed by listening to a series of audio aligned with phonetic-search matches). To ensure objectivity, keyword choices can be selected based on words in the question itself. Difficulty assignments can be based on the number of different paths available, the ranking of the results in the hit list, and how close the hit localized the path within the recording. For example, a question with many search paths leading to a point in the recording within seconds of the answer would fall low on the difficulty scale. Questions with no paths would have high-difficulty.

With memory-retrieval, subjects' memories can influence the search. This can be incorporated into a subject-specific difficulty measure. The pre-questionnaire can be used to assess this. For example the more-accurate remembrance of the time of the event, the place, other speakers, etc., would lower the difficulty.

5.1.11 Salience

An audio recording device does not interpret or bias any sound it receives. It merely receives the sound waves in the air, converts it to an electrical signal, codifies that signal and stores it. It is oblivious to the content and circumstances surrounding it. But, memory aids could benefit from some identification of salient parts. The current approach addresses this to a certain extent. The information-retrieval approach tries to determine salience based on a user-specified query; this is how results are ranked. The single-recording interfaces and visualizations attempt to further focus attention on the relevant.

An alternate is to identify salience prior to any user retrieval efforts. The recordings can be analyzed for the most important moments. StartleCam [48] does this using biometric signals. The audio can be analyzed for particularly animated or quiet parts of discussions; alternately, users can simply mark interesting moments.

The current study did not include such facilities, but could benefit from these. Identification of salient episodes so might help limit archival to selected episodes. A reduced search space is easier to manage, faster to search (both by computer and human). The audio data is intrinsically error-prone; the accurate elimination of low-salience data can lead to fewer false positives and might lead to better ranking of results. However, as stated earlier, missing data—especially data a user remembers they recorded—could lead to fruitless search paths.

5.1.12 Portable memory retrieval

One would assume that people would want to perform retrievals in many of the same situations they are recording. Also, there is evidence suggesting that people remember better when the original context of the desired memory is reconstructed [44]. Finally, a retrieval tool, used *in situ*, could take advantage of such context using, for example, an automatically input-constrained search based on current location and other contextually sensed factors. To better understand some of these issues, I conducted a pilot study with the handheld-computer-based memory-retrieval tool described in Section 3.6.

Lessons learned from the conference-talk study (Section 4.2) influenced the design of this tool. It has many of the features of the personal-computer-based tool (e.g., calendar navigation, audio playback accompanied by transcripts, keyword search, phonetic search, etc.). The main technical differences between the personal computer and handheld PDA platforms have to do with computational processing power, screen size, and input mechanism. How all of these affect memory retrieval are interesting; for this pilot, I focused on the input mechanism. Specifically, can a speech-based mechanism work for query input? Data entry via speech (mediated by a speech recognizer) for a more-constrained wearable calendaring system has shown promising results [129].

As mentioned in Section 3.6, current PDAs lack the computational resources needed for large-vocabulary speech recognition and real-time phonetic searching. For the study, phonetic searches were conducted on a server, and speech recognition was done on a laptop computer; results were transferred in real time to the PDA via a wireless network.

I asked subjects to do a conference-talk memory-retrieval task similar to the evaluation described in Section 4.2. Here, they were only allowed to use the PDA memory-retrieval tool. This includes the ability to speak their search queries (something not available on the personal-computer-based version). Subjects could also use any of the conventional PDA-style text-entry mechanisms (e.g., on-screen keyboard, handwriting recognition).

Four subjects participated. In the desktop-computer-based conference-talk study, searching the calendar and audio playback were common; based on informal observations in this pilot study, these capabilities transferred well to the PDA. Regarding speech input, three subjects opted to use speech-based input of their queries; one preferred the on-screen keyboard. Among subjects who used speech input, it did not take more than three attempts on any given speech input for the recognizer to achieve a satisfactory recognition. In a few cases, subjects would proceed with a misrecognized search query since they anticipated the recognizer could still find the correct match with the remaining correctly recognized words or the recognizer might have misrecognized the speech in the stored audio collection also. Hence, both misrecognitions would lead to a successful match.

Commensurate with a small-N pilot study, the observations here are not conclusive. But, in combination with Wong et al.'s observations [129], they do hint that speech-based input has potential and minimally should not be rejected.

Present computation demands still require a server-based approach, but this will likely change within a few years. PDA screen-size issues did not seem problematic; in the conference-talk study, subjects often found the large personal-computer display useful to visually skim large sections of transcripts to help them localize their audio playback choices. In the pilot, the smaller screen seemed to result in more scrolling. But, subjects were still able to find solutions in minutes.

As devices get smaller, display issues will be become increasingly challenging. As shown in the "speed listening" evaluation (Section 4.3), playing speech synchronized with a transcript is a valuable way to reduce listening time. The PDA has a sufficiently large screen to display a fair amount of text, but mobile phones, watches, etc. are likely to have less space. One method for displaying large amounts of serial information on a small visual apparatus is the "rapid serial visualization presentation" (RSVP) technique [103].

This could be one approach to display transcripts to users while maintaining the speed-listening advantage.

5.1.13 Potential application areas

As stated earlier, this research is aimed at assisting healthy individuals with normal memory failures. My experiences suggest student-advisor and possible supervisee-supervisor communications are one potential area of use. However, a few occupations, by their nature do voluminous recording and have regular need for retrieval: ethnographers and news reporters. The lessons learned from this research can be applied to other disciplines as well.

Work on recording events in lecture halls is intended to give students better means to review course material. There have been numerous, notable efforts to digitally archive meetings (audio, video, presentation slides, written remarks on whiteboards, etc.) and make these available for later review. The often-stated goal for such projects is to produce meeting summaries and to provide a searchable resource for those looking for data within past meetings. Assuming the person was in attendance of the original lecture or meeting, these examples are memory retrieval.

Though the intention of this work is towards memory aid, some aspects benefit to spoken document retrieval or more-general multimedia information retrieval. Most notably, the speed listening technique should extend well to most audio-based retrieval tasks.

This research is aimed at assisting healthy individuals with normal, everyday memory problems. It is worth mentioning the potential value of memory-retrieval aids towards a few commonly known memory dysfunctions. Someone who suffers from retrograde amnesia has lost the ability to recall past long-term memories. Anterograde amnesia sufferers can recall past memories, but cannot form new ones[2]. Both conditions are usually associated with some type of brain trauma and would seem to be the promising candidates for a memory-retrieval aid. For either condition, an archive of ones lifetime can help fill in memory gaps. Assuming other cognitive processes remain unaffected, the ability to simply have records of the gaps *and* search through them may be helpful in a variety of day-to-day situation since the need for assistance is amplified.

Alzheimer's disease is a condition characterized by decreasing mental facility; forgetfulness is one of the first symptoms. Although a memory-retrieval aid might be helpful in the early stages of the disease, especially for long-term memory problems, it does require cognitive processing of information, strategizing about search paths, etc. As the disease progresses, it is not expected that such a tool will help.

Returning to computer-based applications, organizational memory systems attempt to preserve the knowledge and expertise within an organization. Organizations that do a good job with this are robust to personnel loss (due to attrition or absence) and can reduce redundancy. This is an active field of work.

[2] This condition was glamorized in the movie "Memento" released in year 2000.

5.2 Social, legal, and privacy issues

How would I prove I was telling the truth?

- Linda Tripp, on CNN's Larry King Live," February 16, 1999 when asked why she [surreptitiously and illegally] recorded conversations with Monica Lewinsky regarding President Clinton's affair with Ms. Lewinsky. [60]

But if someone phones you in Maryland and wanted to abduct your children or had, and was relaying a ransom note or ransom demand, you would not be able to tape that conversation and use it against the kidnapper. You would go to jail. The kidnapper would go free, because it would be inadmissible, it's against the law.

- Linda Tripp, on CNN's "Larry King Live," February 9, 2001 [61]

Social, legal, and privacy issues are nothing new to technological innovation. A telescope has privacy implications. With audio and video recording devices, privacy concerns typically focus on eavesdropping, surveillance, and telephone wiretapping. As Cate wrote in 1997, "No form of communication other than face-to-face conversation and handwritten, hand-delivered messages escapes the reach of electronic information technologies" [21]. Alas, this too is no longer the case.

With regard to computers and privacy, ubiquitous digital audio recording has not been a particularly active discussion topic. Most privacy discussions focus on computer databases that contain health, financial, shopping, personal, etc. records and how these can be used to infer behavioral patterns via data mining and similar techniques. But misuse of electronic communications is becoming a more active area of scrutiny for privacy advocates. For example, email and instant messages, even if intended as private between the communicating parties, may be archived surreptitiously by an intermediary and these can been subpoenaed in court [2]. Mobile communications devices add a new dimension: tracking your physical location. These data, which are often perceived as personal, are becoming increasingly public. One of the more disconcerting things is the uncertainty regarding who has access to the data and what they doing with it.

Digital eavesdropping concerns can be partially allayed by digital encryption, but access to archived recordings is another issue. Audio archival predates computers, but it is expected that more privacy issues will be raised as foraging technologies improves. Ironically, one of the consequences of the present research is improved analyses and retrieval tools for long-term digital archives of personal audio (i.e., foraging tools); video and other data sources are not far behind.

The intent is to empower individuals with better memory via a detailed personal history, but the recordings also include other people. In effect, the archives include partial histories of the lives of conversation partners. One would expect and hope that permission to record is given with informed consent, without duress, and only for the explicit use as a memory aid. One may also hope that people who record would honor the privacy requirements of all recordees for the lifetime of the recordings. But people change; relationships go awry; those who were once trusted may no longer be. Vigilance in protecting the data may wane. When the recordings are stored on a computer, copying and sharing them is no more difficult than sending an email message or posting on a network-based file-sharing service (e.g., Gnutella, Napster, etc.).

Even with the most trusted conversation partners who doggedly protect their data, there are insufficient legal protections ensuring the privacy of self-maintained personal data archives. It is one thing to have the ability to search through one's own past for a memory trigger, it is quite another thing if someone else (e.g., a zealous attorney, a potential employer, a news reporter, a disgruntled employee) has the same ability with someone else's data.

This section explores these and other issues in more detail. While not a comprehensive account of all social, legal, and privacy issues related to computers and digital data archival, it touches upon a few key points related to personal data archival.

5.2.1 Social issues

5.2.1.1 Ephemeral to persistence

Memory assistance might be a noble goal, but the means prescribed in this dissertation has downsides. Ubiquitous, personal data archival, by its nature, will transform traditionally ephemeral information into a persistent store; this includes both high and low moments in life. Babies' first words, seminal childhood moments, etc. can be captured and relived. But, off-the-cuff remarks, disjointed thoughts, politically incorrect jokes, drug-induced irrational behavior can be preserved too. Keeping in mind that multiple participants in a conversation would all have copies, what are some of the consequences of such archival? Some examples might include:

- "He said, she said" arguments can be settled definitively. While this might help settle the age-old "who was right?" question, detailed review of past statements may be counterproductive if used by emotionally charged couples to settle disagreements. But, under the guidance of a professional counselor, this may have utility.
- Political campaigns will have richer repositories of footage to support and refute candidates' past statements and activities. This is becoming increasingly common in contemporary, high-profile campaigns as more and more archival footage of candidates' younger days are becoming available. This is bound to increase.

These examples might be on the extreme. Even so, for most people, the notion of a past statement, photo, etc. coming back and haunting someone can be disconcerting. Email archival practices are giving some hint to this, but email communications are different than oral; the latter are generally less-crafted and less-polished. Public figures seem to get in more trouble with the press for misspeaking versus miswriting. A few other possibilities:

- Lying may be harder. With a reference source to verify past statements, people might be disinclined to intentionally or unintentionally contradict oneself.
- Personal data archival may encourage people to think more before they speak. The potential for social consequences by saying something foolish may increase if the

utterance is recorded. Also, as grassroots publishing mechanisms like weblogs become increasingly popular, the possibility of publicly distributing quotes from less-public figures increases. A ubiquitous recording apparatus reduces the chance of missing such quotes.

A few times in my experiences of iRemember, I elected to not record a particularly private conversation (either of my own volition or at the behest of my conversation partner). I was the only one who could access the recording, but the mere existence of the recording would have been troubling. In my case, thoughts of computer hackers breaking into the server and copying recordings concerned me from time to time. I took painstaking precautions and never found any evidence of such an intrusion.

5.2.1.2 Awareness

The recording protocol (Appendix A) specified a variety of mechanisms that ensured both subjects and bystanders were aware when recording took place. It was important to not only adhere to these, but to confer with the community of subjects and bystanders to ensure they felt the mechanisms were satisfactory. Assuming recording devices are deployed to the general public, it is unlikely that all users would satisfy the requirements of a university human-subjects review board, the privacy preferences of all recordees, or local laws.

Assuming wider-scale deployment, the recordee should have a mechanism to identify all nearby recording devices and their current state. For example, video cameras typically illuminate a red light when recording so recordees are aware. A lens cap can be used to further ensure recording is not taking place. Ideally, a privacy-concerned recordee could prevent recording devices from collecting data about them. Though I could propose mechanisms for this (e.g., smart cameras that blur out unwilling recordees), it is not clear how to enforce compliance.

Awareness mechanisms for audio are not as commonplace. Microphones can be the size of pinholes and covering one may not fully insulate it from capturing nearby sounds. Parabolic microphones can capture audio from far distances; laser microphones can pick up audio from vibrations on glass. Caution is advisable. However, preventing audio eavesdropping might be possible. A futuristic mechanism might intercept audio from the larynx, digitize it, and transmit the data encrypted to only the desired recipients.

For the time being, awareness rests with the recorder, their understanding of local laws, privacy, and ethics.

5.2.1.3 Data access and protection

Once recording mechanisms are in place, some questions arise: where are the recordings stored? who has access to them? what is being done to protect them?

For the current studies, the data were stored on a computer server connected to the Internet. The server-based approach is needed due to data storage and CPU processing requirements. This is likely to change within a few years. The server-based approach also simplifies the implementations for clients across different platforms and devices (i.e., personal-computer-based, handheld computer). The server-based approach has the unfortunate consequence that "hackers" can try to intercept audio transmissions or break into the server to access the recordings.

The digitized audio recordings were transmitted over the computer network, but were encrypted using the Secure Socket Layer (SSL) protocol. This type of data transmission is widely regarded as a secure means of digital communications. The recordings were not encrypted on the server, but this could have been easily added and doing so is advisable for future implementations.

I had access to the recordings, including those recorded by others. This was helpful for the purpose of doing studies, but not so for any other reason. Ideally, access should be restricted to only the recordees of a given conversation. While these mechanisms may be helpful to thwart voyeurs and hackers, the discussion on legal issues addresses how law-enforcement officials can compel the release of such recordings.

5.2.2 Legal issues

The recording laws do not favor those who record. Most states in the U.S.A. have laws requiring consent before initiating audio recording. There is some state-to-state variation with respect to both the setting (public versus private) and how many people must consent (one or all) [51]. Behind all of these laws is the basic premise asserted by the Forth amendment of the US constitution:

> *The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.*

Though it may sound protective, in practice, it falls short of protecting personal audio recordings and similar self-made records, regardless of the purpose. The right guarantees a person's right against government-directed search and seizure of private property, but is tempered by phrases like "unreasonable searches" and "probable cause." "In the interest of justice" and "in the public's interest" are also phrases that litter various other legal writings. From a legal perspective, a close cousin to a memory prosthesis is a personal diary. U.S. courts have addressed protection of personal diaries, and their current position is diaries can be searched and seized [53]. Hence, it is unlikely that the less-private memory prosthesis would be afforded more protection.

The most relevant legislation in the US is the 1986 Electronic Communications Privacy Act (ECPA) [110]. Telephone wiretapping statutes were already in place and ECPA updated these to address computer-based communications; its intention is to protect individuals' electronic communications. For example, it stipulates penalties for anyone who "intentionally intercepts, endeavors to intercept, or procures any other person to intercept or endeavor to intercept, any wire, oral or electronic communication." This includes the government, employers, voyeurs, etc. The European Union has similar statutes regarding electronic communications (EU Directive 2002/58/EC [35]).

The US Patriot Act of 2001 [111] amended the ECPA to allow additional provisions for government access to communications (e.g., only a search warrant is needed to access stored voicemails). Another Patriot Act amendment allows organizations (e.g., internet service providers) to voluntarily release stored electronic communications under emergency situations without permission of the communicating parties.

The laws relevant to the communications are of present interest so long as memory-aid-related data are transmitted electronically through the telecommunications infrastructure. With a non-server-based approach, some of the troubling aspects of these laws might not apply. However, the storage aspects would. For example, if an internet service provider hosted the computer and hard disk in which the personal data archive was stored, there would be many legal means for government officials to access the data. However, if an individual kept all data on a portable computer, there would be fewer legal means to access the data.

Hiding data is an option and the safest place might be inside one's body. While the courts have not set a limit to what can be seized, the standard for extracting things from one's body is higher than things outside [88]. Automobile drivers can be compelled to give breath, urine, or blood samples. A court can order a person to submit blood samples for DNA testing. As one goes further into the body, the courts have shown restraint. It is unclear if a person can be forced to have an implanted data storage unit surgically removed for evidence gathering.

Legal issues surrounding personal data archival are not limited to surveillance, eavesdropping, and search warrants. Having a record of one's past can have legal benefits. For example, just as the recordings can be used to incriminate, they can be used to exonerate. Loftus has studied false memories and implanted memories and the dangerous impact it has on eye-witness testimony [70]. A detailed record of victims', perpetrators' and witnesses' circumstances surrounding a crime can lead to a more-accurate reconstruction of the events. A wrongly accused defendant would have the necessary evidence to deflect an incorrect prosecution. It may be in one's interest to document one's life in detail.

5.3 Future work

Current research has only scratched the surface of memory assistances afforded by ubiquitous computing. Lessons learned from iRemember, while promising, are limited to a small set of memory assistances and a restricted population. A rich future is in store.

In this section, proposals for future memory aid research are presented. These attempt to address additional memory problems, use additional data sources, reduce the social and legal barriers, and address memory problems among the elderly.

5.3.1 Memory-encoding aids

Instead of remedying a memory failure, can a memory aid reduce the chance of the failure from happening in the first place by improving biological memory-encoding? One theory suggests humans encode memories in summary form. Details are filled in as more time, effort, and attention are spent on a particular event. The more details one encodes, the more association paths will there be to the memory. The amount of detail one remembers about a particular event depends both on one's ability to find a stored memory and how much effort is dedicated to the initial encoding process. The "depth-of-processing" theory suggests that more processing done during memory encoding leaves a stronger memory trace for future retrieval [5,24].

All of this suggests that memory-strengthening efforts soon after an event may mitigate the need for future memory assistance. Computer-based memory aids can potentially facilitate this biological improvement by encouraging well-focused memory-strengthening exercises at the time of biological memory encoding. As mentioned earlier, the applicability of the current memory-retrieval aid is limited to transience and blocking problems in which people both recognize the problem is occurring and wish to remedy it. This limitation is not expected for memory-encoding aids.

Ebbinghaus' forgetting curve [30] shows that people forget over 50% after the first hour and 66% after the first day. This suggests a memory-encoding aid would be most effective within the first 24 hours of an event. In addition to formation of new memories, maintaining existing memories can also benefit from periodic rehearsal. A single test of a previous memory, even after years, has significant impact on its subsequent retrievability [67]. This suggests a memory aid that occasionally elucidated past relevant events (even if only once every few years) would help improve retention of such events.

Despite its potential benefits, computer-assisted encoding assistance might interfere with the normal, biological selection of memorable events. Forgetting not only has to do with the amount of time that passes between event and retrieval, but how much activity occurs between event and retrieval [6]. Hence, it could be argued that encouraging memory-strengthening exercises have the adverse effect of artificially amplifying the strength of computer-selected salient memories in lieu of an individual's biological selection process. It remains to be seen if this selection bias will have such an effect among the subjects.

One vision of a memory-encoding aid is as follows: as events transpire in one's life, the aid would try to identify moments most worthy of later remembrance, record those bits, and before one's biological memory has faded, present a summary of the salient parts.

This presentation would hopefully inspire rehearsal and consequently, help strengthen the biological memory of the event. This ideal is very difficult to achieve. In fact, this vision is loaded with challenging sub-problems. Two that stick out include the identification of salient events and subsequent summarization. Interestingly and perhaps conveniently, these are also challenges in the design of the memory-retrieval aid.

5.3.2 Limited archival

It is worth considering the tradeoff between remedying potential memory problems that require long-term archival and the benefits of limiting archival. The reported scenarios in which the memory prostheses were mostly put to use (i.e., reviewing student-advisor conversations for a pending deadline) can succeed without long-term archival. Task completion may be an opportune time for data purge, or at least, extraction of salient parts and deletion of the remains. A memory-encoding aid intrinsically does not require long-term archival. The necessity for persistence implied by a lifetime archive, at least for certain memory assistances, is not clear.

A limited- or non-archival strategy has additional benefits. First, a restricted search space may improve search experiences by reducing the time to find answers (or abandon fruitless search paths). Second, an illicit data intrusion would have limited ramifications. Third, if a user were embroiled in legal struggles in which recordings were subpoenaed, an established deletion strategy may avoid allegations of illegal destruction of evidence. Finally, conversation partners might be more willing to be recorded if there is an agreed-upon destruction policy.

5.3.3 Non-audio data

The focus of our current prototypes are on audio as a memory-triggering resource. However, with the data-collection infrastructure in place, adding new data sources is now a straightforward process. *Doppelgänger* [83] is an open architecture for adding heterogeneous sensors to user-modeling systems. A similar approach can be used for gathering resources for memory triggering.

Visual

Previous work at the Media Lab to build technology to remedy "visual amnesia" [74]—a natural addition to the current system—provided users with a visual-capture capability. Not surprisingly, among those who have tried iRemember for short periods of time (one to five days) the most commonly requested capability is visual recording. There is an active research community investigating general-purpose video retrieval [102] and it is expected that the mere existence of a visual triggering source—even without any content analysis—will act as a valuable memory triggering resource. Formal evaluations would be needed to better understand the impact of visual triggering and if image quality and image content analysis play a significant role.

Biometrics

Biometrics may serve as a vehicle to identify instances of salient events. Devices for continuous biometric recording for personal health monitoring have been studied [43] and commercially available biometric sensors are becoming readily available. While the technical hurdles for continuous biometric recording are ceasing to be challenges, there is limited evidence relating memory to biometrics. What little is known is related to memorability of fear response in animals and the role of affect in the formation of "flashbulb memories."

The fear studies are similar to Pavlov's classical conditioning experiments. LeDoux showed that when presented with a fearful stimulus, animals quickly learn the conditions under which the stimulus occurs [62]. In terms of biometrics, blood pressure, galvanic skin response, and physical agitation are perturbed by presentation of the fearful stimulus. If, during training, the fearful stimulus is presented in conjunction with a conditioned stimulus (e.g., an audio tone, a light, etc.), the same physiological reaction will result when the animal is presented with only the conditioned stimulus. In effect, these studies have quantified the intuitive notion that a memory of a fearful experience can elucidate a biometric and emotional reaction without actually reenacting the experience.

Flashbulb memories refer to memories of events of such significance that observers can recall their circumstances (who, what, where, when, emotional response, etc.) with abnormally high detail, clarity, and durability [39]. These are often associated with elevated emotional states and many people claim to experience this phenomenon with hearing news of personal or public landmark events (e.g., death in the family, JFK assassination, Space Shuttle Challenger disaster, September 11 terrorist attacks). Theories on how these memories are formed differ in the mechanisms involved; yet they are consistent in that all suggest a relationship between elevated affective state and memory.

This limited evidence leaves a thirst for more data regarding *any* possible relationship between biometrics and human memory. While it is premature to assert much about this interaction, biometric sensor technology has reached the point in which hypothesis-driven, controlled experiments can be conducted. Such work is needed to clarify this murky picture.

5.3.4 Memory and aging

This research project focuses on healthy people in academic and knowledge-worker settings. But, there is potential for impact on a much broader range of individuals, groups, and settings. Normal, healthy individuals suffer decreased memory facility due to aging and United States demographics suggests increasing numbers of the "baby boomer" generation are reaching ages where decreased memory facility has more pronounced effects on daily life. The elderly feel a great deal of anxiety as a result of memory loss and that anxiety is a good predictor for future cognitive decline [101]. While there may not be medicinal means to prevent the cognitive decline, having better memory aids may relieve some anxiety. In fact, the elderly, while showing decreased performance in laboratory memory tests compared to their younger counterparts, often perform adequately if not better on daily life tasks due to more efficient and diligent use of memory aids [10]. There may be near-term potential for improved quality of life for larger segments of an aging population through better and more accessible memory aids. Related to these goals, an exploration studying family histories illustrated how archived news media can help trigger memories among the elderly [46]. More personal, family-oriented archives may be of similar value.

5.3.5 Infrastructure-assisted capture

Public and semi-public environments are already filled with devices recording a variety of day-to-day activity. Buildings, roadways, and automatic bank teller machines are often equipped with video surveillance and these have proven useful in solving crimes. In Australia, taxicabs have continuously recording cameras (both interior and exterior); some police cars in the United States are similarly equipped. Broadcast and sports media outlets vigilantly document publicly witnessed newsworthy events. It is becoming commonplace for universities to broadcast and archive lectures, seminars, and classroom discussions through "distance learning" programs. Similar video-conferencing systems are used for more-private communications in a variety of settings. Use of personal digital cameras (still and motion) has risen sharply over the past few years, especially via mobile phones. Many web-based cameras provide continuous feeds of public locations. We and the events we witness are recorded far more often than we think [15]. Coverage is bound to increase.

The responsibility of memory-aid recording need not rest solely in the hands of each individual. The previous examples illustrate existing infrastructure already capturing a valuable and largely unused memory-triggering source. The technical hurdles to integrate with the existing memory prosthesis are not large. For example, a user control to identify interesting moments along with a mechanism to collect or tag data from nearby recording apparatuses would suffice. As with the memory prosthesis, the engineering hurdles are overshadowed by the social, ethical, and legal ones. Less-restricted access to such surveillance data may draw ire; Brin discusses some scenarios in detail along with cautions against hypocrisy on data-sharing stances [15]. Projects allowing people to avoid public-surveillance might avert some controversy [63].

Chapter 6 Conclusion

> *Here's something that happened today. I was in a meeting and the door was open to the Weisner room so I was slightly in earshot for Felice. ... I was with Cynthia, I was struggling with the name of somebody from AARP and I remembered Horace Deets, but I couldn't remember his title; and I remembered our liaison, but I couldn't remember her name. Felice then says, Dawn Sweeney. She was acting as a memory prosthesis. ... That is the goal.*
>
> - Walter Bender

Computers pale in comparison to an able assistant; not all of us are as lucky as Mr. Bender. If computers can come close to this example above, we are well on our way to reducing some of the more-serious and damaging memory problems in our lives. The memory-retrieval approach presented herein illustrates one way of achieving this.

In particular, this dissertation demonstrated how long-term personal-data archives can be used for memory assistance. The technical hurdles of accruing such archives are quickly diminishing; the increasing interest in personal data archival suggests increasing numbers are on the verge of engaging in such endeavors. Turning such collections into a memory aid is the challenge. The results of the evaluations illustrate that the use of information retrieval technologies can be a viable course.

The approach is probably not ready for widespread use, but it may not be long before it is. The subjects in the evaluation (especially the personal-experiences study) are adept at using research prototypes; the general populace is probably somewhat more finicky. Speech recognition accuracy is still low for spontaneous speech in the noisy environments typical of everyday life; but, if current progress is any indication, improvements are likely which, in turn, should benefit memory retrieval. The visualizations and spoken document retrieval techniques also have room for improvement. The current implementations are rough prototypes and redesigning with additional care for human-factors and performance considerations would likely result in improved user performance.

Addressing transience and blocking problems were the initial goal. In the evaluations transience memory problems were dominant; although some instances of blocking were observed in the conference-talk evaluation, the examples were too few.

The evaluations also revealed some challenging problems. First, misattribution memory problems were not uncommon, but it is unclear how the memory-retrieval approach can address these. A user's false belief in a memory can lead them astray in their search paths; it is not clear how the computer can help identify these problems on behalf of the user.

The evaluations revealed that people with poor remembrance of specific episodes are disadvantaged by the reduced ability to find specific episodes. This poor remembrance does not necessarily imply a bad memory, it may simply mean the person is inundated with many similar episodes in their lives such that specific episodic details are blurred by overload. In theory, one reason to record one's life in such detail is to minimize the need to remember the details. Remembrance of certain details were useful; perhaps the archival tool will help people focus only on the details most useful for future retrieval (who, where, when, etc.) and let the computer capture the rest (what, why, etc.).

I conclude by listing the contributions and adding a few final thoughts.

6.1 Contributions

The main contributions of this thesis relate to individual memory assistance. However, the applications of these are not necessarily limited to this domain. Below is a list of the main contributions.

- Methods for conducting long-term personal-experience-based memory-retrieval experiments. Methods for conducting traditional memory experiments are well-established. There is a rich history and broad shoulders to stand upon. But, conducting memory-retrieval experiments (especially free-form) is relatively new. Furthermore, the challenge of conducting such tests with personal experiences over the long-term invariably restricts investigators to a small subject pool; the demands on these subjects are high and experimental-design mistakes can be costly. The value of vetted experimental designs can be further illustrated by the increasing interest in computer-assisted personal-data archival and the potential utility of these archives (and a larger subject pool) for the next round of memory-retrieval experiments.
- Evidence that people can do memory retrieval for transience memory problems within minutes. In both the conference-talk and personal-experience evaluations, it took subjects minutes to find answers. This is far short of re-listening to the entire archive and adds credence to the approach. Each task and situation will determine if this is a small enough commitment to consider an attempt. This is also an early attempt at the problem and subsequent iterations will likely benefit from improved user interfaces, speech recognition, and retrieval techniques. Hopefully, the time measures here are closer to the upper bound.
- Identified technologies where resources should be focused. For the memory-retrieval task, phonetic searching is a technology worth improving; sophisticated landmark-based searching can be left as is. Better microphoning for daily-life situations can also help. Improved speech recognition is also a clear choice, but that is related more towards general spoken document retrieval and not memory-retrieval in particular.
- Development of the "speed listening" technique that can reduce audio-listening time while maintaining comprehension. Though inspired by the present audio-based memory-retrieval problem, the application of this extends to other audio-listening situations.
- Identified a potential application area: supervisee-supervisor communications, especially when task-based and deadline-impending. The evidence presented is based on anecdotal and self-reported student-advisor communications.

6.2 Final thoughts

Weiser describes ubiquitous computing as those that "weave themselves into the fabric of everyday life until they are indistinguishable from it" [120]. The iRemember prototype is bulky by today's standards and probably would not satisfy Weiser's criteria; a smaller always-available memory aid that integrated smoothly might do so. The memory assistance capabilities might satisfy part of Vanevar Bush's Memex vision.

The value of having better memory aids need not be limited to improved meetings, performing better on tests, or any objective metrics. Prof. Elizabeth Loftus, her students, and I were discussing the intrinsic confounds of personal-experience-based memory-retrieval evaluations. One clever metric they suggested was to see how other people perceived the wearer's memory. After some period of day-to-day usage, if others perceive the wearer as someone with a better memory compared to when they did not have the aid, the memory aid has successfully conveyed social value (even if there is no empirical evidence supporting improved memory). This could be analogous to the perception that people wearing glasses look smarter than those that do not.

Occasionally, I'm asked what are the major hurdles that need to be crossed before a memory aid like iRemember is available in the marketplace. I feel that the social, legal, and privacy issues are more challenging than the engineering ones. Long-term archival of personal data can rub against our sense of privacy. The lack of legal protections can threaten the ownership of the data. This is not to say that these issues are insurmountable; there are societies that may be ready for this sooner than others.

Appendix A : COUHES application

The research in this dissertation involved human subjects. Accordingly, experimental protocols were submitted to and approved by M.I.T.'s Committed on the use of Humans as Experimental Subjects (COUHES). Most of these protocols were unremarkable. However, the application submitted as part of the study described in Section 4.4 stood out due to the many privacy invading possibilities and possible snags in protocol compliance. The substantive part of the application is included below.

A.1 Purpose of study

We propose a set of experiments designed to explore the nature of human memory recall using audio and contextual cues from past events. For these experiments, we wish to break away from the confines of laboratory human memory tests and focus on real world situations. Previous research on real-world long-term human memory recall was based on physical contextual cues and not audio [67,116].

In contrast to a previous experimental protocol approved by COUHES (#2701), this protocol will extend outside of a conference room and into the natural real-world setting. Subjects will be given a portable recording device to collect data through the course of their normal day-to-day happenings.

In broad terms, these experiments will consist of recording informal discussions during any of the variety of conversations that take place during a typical day.

A.2 Experimental Design

Two categories of experimental subjects

Before describing the details of the experiment, we note that there will be two categories of subjects in these experiments. The first category will be a small set of people (between 5–10 people) who volunteer to carry and/or wear recording devices with them. Henceforth, these subjects will be called "prosthesis wearers."

A good portion of the data recoded will pertain only to prosthesis wearers. However, one type of data collection will involve audio recording of conversations. Some of the people audio recorded will not be prosthesis wearers. They will simply be associates of prosthesis wearers who agree to be audio recorded. Those people who volunteer to be audio recorded by prosthesis wearers, but are not prosthesis wearers themselves will henceforth be called "non-prosthesis wearers." We expect approximately 20 non-prosthesis wearers per prosthesis wearer.

There will be two different informed consent agreements. Subjects wishing to participate in the experiments will be asked to read and sign the agreement appropriate to their participation.

Basic recording system

Prosthesis wearers will be asked to carry a portable recording device with them. Henceforth, this device will be called the "wearable memory prosthesis." The wearable memory prosthesis is a palm-sized computer that will be capable of recording audio, recording the location of itself, recording biometric data from the prosthesis wearer, and recording a log of the prosthesis wearer's interactions with the device. Each of these recording features can be turned on or off separately depending upon what data each prosthesis wearer wishes to volunteer. Figure 1 shows a photo of the wearable memory prosthesis. As the data are recorded, they are transferred to a separate computer that is designed to securely store the data (data security will be discussed later).

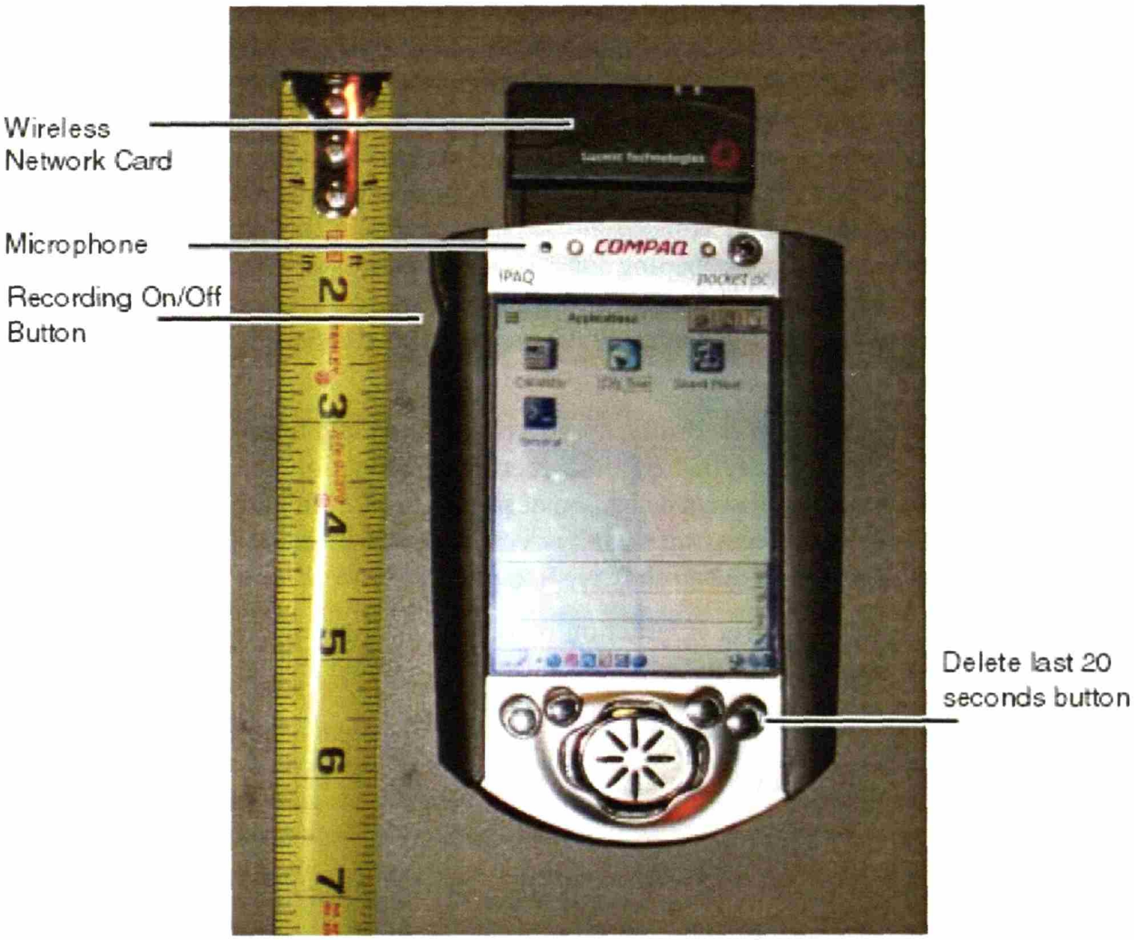


Figure A-1: The wearable memory prosthesis

Audio recording

The audio recording system is designed to be fairly simple. Pressing the button on the left of the device (shown in Figure) activates and deactivates recording. Separate, unique, audible tones are played when the recording is activated and deactivated.

Location recording

When outdoors, the device can identify it's location using the Global Positioning System (GPS). When indoors, the device can identify it's location using the 802.11b wireless computer network. For example, inside the Media Lab, there are over a dozen stationary 802.11b wireless access "base stations." A computers with a wireless computer network card can determine it's approximate location by computing the signal strength to nearby stationary "base stations" and then employing a "triangulation" technique. The wearable memory prosthesis will use such a wireless network card (shown in Figure A-1) and attempt to determine its location.

Biometric recording

Several biometrics sensors will be used to collect information about the prosthesis wearer. The proposed set of sensors will measure: heart rate, pulse, skin conductivity (also known as galvanic skin response), EKG, and respiration.

User interaction recording

Along with the recording system, prosthesis wearers will be given computer software designed to help them browse and search archives of recorded data in order to help recall forgotten memories. The investigators wish to know how useful these software tools are and how to design better tools. Hence, the software will be designed to record how subjects use them. This includes keeping logs of what features of the software are used, when, how often, etc. Additional items that might be captured include the queries used during searches. Essentially, any mouse click, menu item selection, typed text, etc. may be captured.

As mentioned before, the audio recording system is activated and deactivated by the press of a button. The other three types of data recording (location, biometric, and user interaction) can be turned on always or can be designed to turn on and off along with the audio. Each prosthesis wearer can individually decide how they would like their wearable memory prosthesis to work and when they wish to give data.

The location, biometric, and user interaction recording systems are recording data only about the prosthesis wearer. The audio recording system will be able to capture data from both the prosthesis wearer and nearby non-prosthesis wearers. This is an important distinction in terms of who must be informed when different recording systems are activated. This is discussed in detail in the "Awareness" section below.

Awareness

It is important that all subjects always know when the wearable memory prosthesis is recording data. Since recording audio data has special legal and privacy concerns versus the other types of recorded data, audio recording will be discussed first.

Audio

Before a prosthesis wearer begins audio recording, they must:

1. Insure that all potentially audible nearby persons are part of the experiment by virtue of having signed an informed consent agreement. Prosthesis wearers will be told who among their associates have signed informed consent agreements allowing audio recording.

2. Verbally obtain permission to audio record these persons for that particular instance

It is desirable that prosthetic wearers audio record this verbal consent using the wearable memory prosthesis, but it is not required. An attached Appendix describes Massachusetts state law pertaining to audio recordings.

As mentioned before, when the recording button is pressed, an audible tone is played indicating that recording has begun. When the prosthesis wearer presses the button again to stop recording, a different audible tone is played indicating that recording has stopped. Furthermore, the display on the wearable memory prosthesis will show a visible sign indicating when the audio recording system is on. This sign will be visible and legible to anyone within several feet of the wearable memory prosthesis.

"Nearby persons" and discernability

Massachusetts laws do not define strict boundaries and our use of the phrase "nearby persons" is also a loose definition. The key issue here is how close or how loud must a person be to the wearable memory prosthetic to be considered a "nearby person" and hence informed of an active audio recording device. We have conducted tests on the microphone of the wearable memory prosthesis and found that it is optimized for short distance recording: typically within 2-3 feet. Nevertheless, the microphone can still pick up conversation level sounds at greater distances. The point when the audible quality of the recording ceases to be discernable is a conversation taking place 10 feet from the wearable memory prosthesis.

However, a 10-foot rule is also insufficient since it does not take into consideration louder speakers. What is needed is a simple metric for prosthesis wearers to use so they know when someone should be informed that the wearable memory prosthesis is recording. The metric we wish to employ is discernability. If a conversation is discernable to the prosthetic wearer, then participants in that conversation must be informed or the audio recording system should not be activated. If only mumbles can be discerned from a conversation, then those people need not be informed that an audio recording is taking place. Our tests suggest the wearable memory prosthesis' microphone is definitely not better than normal human hearing. Hence, this metric should be simple enough for prosthesis wearers.

Even with this, we are relying on the subjective opinion of a person to determine if a nearby conversation is discernable. Discernability can be affected by fatigue, attention, "cocktail party effect," etc. Hence, prosthesis wearers will be asked not to activate audio recording if there are many simultaneous nearby conversations or if they are particularly tired or under the influence of any drug/substance that may impede their ability to determine discernability.

Inadvertent recordings

Recordings may happen in offices, meeting rooms, hallways, near the water cooler, etc. Though a recording may start with all nearby persons informed, others may walk within recording distance before a prosthetic wearer notices or has time to deactivate the recording. Many precautions can be taken to prevent a recording of someone who was not informed, but the sheer diversity and uncontrollability of a natural real-world setting suggests that accidents will happen.

If this happens, prosthetic wearers can press a button on the wearable memory prosthesis to delete the previous 20 seconds of audio recording. This button is shown in figure 1.

If more than 20 seconds have elapsed since the inadvertent recording, the prosthetic wearer will be asked to delete the entire recording or they can ask the investigators to delete it for them.

Even with this precaution, the investigators will periodically listen to random samples of audio recordings to ensure full compliance with experimental protocols and all informed consent agreements. Any non-compliant recording discovered by the investigators will be deleted.

Location, biometric, and user interaction

With regard to location, biometric, and user interaction recording, only the prosthesis wearer is providing data about themselves. Hence, no one else needs to be informed when these data recording systems are active.

As mentioned before, each prosthesis wearer can individually decide if they want these recording systems to be activated only when the audio recording is on, or if they want these three systems to be activated always. When a prosthesis wearer declares that they wish to have the prosthetic device recording certain types of data always, it is assumed that the subject will know that whenever they are using or carrying the device, it will be recording. When a prosthesis wearer wishes to enable location, biometric, and user interaction data recording only with audio recording, the aforementioned audible tones will play normally.

Non-prosthesis wearers and location

In the previous section, it was stated that no one needs to be informed when the location recording system is active. This is not exactly true. There is one situation in which the location of a non-prosthesis wearer can be determined. If the prosthesis wearer enables both audio and location recording, the location of anyone who is audio recorded by the wearable memory prosthesis can be deduced. The informed consent agreement for non-prosthesis wearers reflects this.

A.3 Retrieving recordings

This study is investigating how memory prosthetics are used to aid in everyday real-world memory problems. Hence, prosthesis wearers will be given computer software tools designed to help them quickly and efficiently search for, browse, view, and play recorded data. Prosthesis wearers will be given this software and it will be restricted to allow them access to only their recorded data. The investigators will have access to all recorded data from all subjects.

A.4 Security and ultimate disposition of data

Data collected by these experiments are intrinsically private. Technological constraints require that the data be stored on a central computer system with large data storage capabilities. It is imperative that this central storage location be access restricted. The investigators will be taking many precautions to ensure the security of the data.

First, the central computer will be physically secured. In addition to being in a locked office, keyboard and mouse access to the computer will be disabled through the use of a physical key. Only the investigators will have a copy of this key. Second, though this central computer will be on the computer network, it will be password protected and only the investigators will have the password to the machine. Third, despite password protection, there is always the threat of computer "hackers". To protect against this, network traffic to and from the machine will be restricted to within the Media Lab. The computer network technical support team at the Media Lab's can configure a computer's network connection to be "blocked." in this way.

Who has access

On the central computer system, data from all subjects will be stored. A password security system will be employed to ensure each prosthesis wearer will have access to only their data. Non-prosthesis wearers will not have access to any data (including recordings of their own voice). Naturally, the investigators will have access to all of the data. With respect to investigator access, some people who have expressed interest in being a subject have also expressed concerns about too many investigators having access to their data. Each investigator on the project who may have access to the data will be individually listed on the informed consent form and each subject can selectively decide whom they wish to allow access.

Naturally, subjects who request too many restrictions with regard to investigator access may not be able to participate in the experiment.

Ultimate disposition of data

We expect these experiments to take approximately one year. In order to have time to publish results from the experiments, the data will be kept securely for two years after the conclusion of the experiments. After that time, unless separate written authorization is given by all recorded subjects (including non-prosthesis wearers), the audio data will be destroyed.

With regards to location, biometric, and user interaction data, all of this will also be destroyed two years after the experiments. However, if prosthesis wearer wishes, they can voluntarily contribute their data beyond the first year thorough a separate written agreement.

A.5 Experimental subject commitment and acquisition

Prosthesis wearers

Subjects wishing to participate as prosthesis wearers will be asked to carry and/or wear the wearable memory prosthesis for 1-2 months. They will be asked to wear the device whenever convenient. A suggestion of 2-3 hours per day will be given, but the prosthesis wearers will not be bound to this. Furthermore, prosthesis wearers will not be required to record data at any specific time nor will they be given quotas. Each prosthesis wearer can selectively choose when and what they wish to record.

After the 1-2 month wearing period, prosthesis wearers will continue to have access to the software tools that allow them to search and browse their data. Prosthesis wearers will be allowed access to these tools for 6 months to 1 year after they finish the "wearing" period. These subjects will not be required to use these tools.

Prosthesis wearers will be asked to participate in periodic memory tests. This will include both short questionnaires and short interviews. These tests are expected to last about 20 minutes and will occur approximately once per month for the duration of the experiment. Furthermore, prosthesis wearers will be asked to fill out surveys and participate in interviews to describe their qualitative experiences of using the prosthesis both in terms of the affect on their memory needs and the acceptability of the device in their social environments. These tests also will occur approximately once a month. In total, we expect about 2 tests/surveys/questionnaires per month for 20 minutes each.

Subjects will be drawn from volunteers inside the Media Lab. It is not anticipated that explicit advertising will be needed for this as many people inside the Media Lab are aware of the project and have expressed interest in participating as a prosthesis wearer.

Non-prosthesis wearers

Prosthesis wearers will be asked to identify people with whom they commonly interact. These people will be approached and asked to participate in as non-prosthesis wearers. The commitment for these subjects includes allowing audio and location recording whenever they feel comfortable. Given prosthesis wearers must get verbal consent before each recording, a non-prosthesis wearer can always decline to participate in any individual recording.

Additional requested commitments for non-prosthesis wearers including providing a sample of their voice. This sample will be analyzed and used to develop a speaker-identification system that can attempt to determine who is speaking during a given audio recording. Furthermore, non-prosthesis wearers will be asked to participate in questionnaires and interviews. These will include questions about their experiences with the prosthesis wearer. We expect 1-2 20-minute questionnaires/interviews per month for approximately 1 year.

Appendix B : Sample questions for evaluation 1

Questions used as part of the conference-talk evaluation in Section 4.2.

- What side of the street do people drive on in Dublin, Ireland?
- What cities were *specifically* mentioned in which the BeatBugs had been tried as part of workshops or concerts? Briefly describe, in your own words, what a BeatBug sounds like.
- What group is working on the "actuated workbench?"
- Which projects or people *specifically* cite the use of a speech recognizer?
- What book (title and author) does Mitchel Resnick cite?
- Worldwide, how many people are using Open Mind? How may facts have been entered into it?
- On average, how many emails do Media Lab sponsors get per day?
- According to Andy Lippman, what is missing from all students' presentations?
- What "gets smarter for you, through its interaction with you?" Who said so?
- Do people tend to buy more with a smaller or larger selection of choices?
- What sensors were mentioned as part of the tabletop robot in Deb Roy's group? What is the name of the robot?
- Who are the key people responsible for the Communications Futures Program?
- What is the name of the robot in Cynthia Breazeal's group? Who built it? How many people would it take to control such a robot?
- How big is OMCSNet (number of concepts & links)?
- Bruce Blumberg describes two hard problems in Andrew Brooks' research. What are these?
- How is "relational agents" defined?
- A project was cited that helps monitor environmental hazards. Please name any/all of the following: the project name, the person working on the project, the types of environmental hazards that can be monitored, where the system will first be deployed.
- Who did Nathan Eagle specifically mention he is collaborating with?
- What was a specific, cited example of a bad decision making process from recent events?
- Who is working on "spatial computation" and what it?
- What is more fun than a box full of hamsters?
- What two factors are cited to impact group decision making?
- Within the first two years of the internet coming out, what percentage of bits went towards social interaction?
- Which company provided some data to the Media Lab, but did not give permission to present results in time for the DL meeting?
- One speaker described how people change decisions making based on different emotional states. Which emotional state was described? Who was the speaker?
- To what book did Hugo Liu apply his common sense emotion detection algorithm?
- What did David Reed talk about two years ago at the Digital Life meeting?

Bibliography

1. Abowd, G.D. Classroom 2000: An Experiment with the Instrumentation of a Living Educational Environment. IBM Systems Journal, 38(4), 508–530, (1999).
2. Arons, B. Techniques, Perception, and Applications of Time-Compressed Speech. *Proc. 1992 Conference, American Voice I/O Society*, 169–177. (1992)
3. Arons, B. SpeechSkimmer: Interactively Skimming Recorded Speech. *Proc. UIST 1993*, 187–196. (1993).
4. Associated Press, Think before you text. http://www.cnn.com/2004/TECH/ptech/06/07/text.messaging.records.ap/ (June 7, 2004).
5. Baddeley, A.D. *Essentials of Human Memory*. Psychology Press Ltd. (1999).
6. Baddeley, A.D., Hitch, G.J. Recency Re-examined. In S. Dornic (Ed.), *Attention and Performance,* 647–667. Hillsdale, NJ: Lawrence Erlbaum Associates Inc. (1977).
7. Bahl, P. and Padmanabhan, V.N., RADAR: An In-Building RF-Based User Location and Tracking System. In *Proceedings of IEEE Infocom*, Tel-Aviv, Israel (March 2000).
8. Bahrick, H.P., Bahrick, P.O., and Wittlinger, R.P., Fifty years of memory for names and faces: A cross-sectional approach. In *Journal of Experimental Psychology: General*, **104**, 54–75. (1975).
9. Bahrick, H.P. and Phelps, E. Retention of Spanish vocabulary over eight years. In *Journal of Experimental Psychology: General*, **104**, 54–75. (1987).
10. Balota, D.A., Dolan, P.O., Ducheck, J.M. Memory Changes in Healthy Older Adults in *The Oxford Handbook of Memory* (eds. Tulving, E. and Craik, F.I.M.) Oxford University Press. 395–409. (2000).
11. Beasley, D.S. and Maki, J.E. Time- and Frequency-Altered Speech. In N.J. Lass, editor, *Contemporary Issues in Experimental Phonetics*, Academic Press, 419–458. (1976).
12. Blaney, P.H. Affect and Memory: A review. Psychological Bulletin, 99, 229–246. (1986).
13. Borovoy, R.D., Graves, M.J., Machiraju, N.R., Vemuri, S. System for automatically retrieving information relevant to text being authored. U.S. Patent 5,873,107. (1996).
14. Bove, V.M. and Sierra, W. Personal Projection, or How to Put a Large Screen in a Small Device. In *Proceedings of SID* (2003).
15. Brin, D. *Transparent Society*. Addison-Wesley (1998).
16. Brown, R. and McNeill, D. The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, **5**, 325–337. (1966).
17. Budzik, J. and Hammond, K., Watson: Anticipating and Contextualizing Information Needs. In *Proc. of 62nd Annual Meeting of the American Society for Information Science*. (1999).
18. Bush, V. As We May Think. *Atlantic Monthly* **76**(1), 101–108. (July 1945).
19. Campbell Jr., J.P. Speaker Recognition: A Tutorial. Proc. IEEE. **85**. 1436–1462. (September 1997).
20. Castro, P., Chiu, P., Kremenek, T., Muntz, R. A Probabilistic Location Service for Wireless Network Environments. In *Proc. Ubicomp 2001*, Atlanta, GA, (September 2001).
21. Cate, F.H. *Privacy in the Information Age*. Brookings Institution Press. (1997).
22. Chellappa, R., Wilson, C.L., Sirohey, S. Human and Machine Recognition of Faces: A Survey. In *Proceedings of the IEEE*, **83**(5) 705–740. (1995).
23. CMU Pronouncing Dictionary. cmudict0.6d. http://www.speech.cs.cmu.edu/cgi-bin/cmudict

24. Craik, F.I.M. and Lockert, R.S. Levels of processing: A framework for memory research. In *Journal of Verbal Learning and Verbal Behavior*, **11**, 671–684. (1972).

25. Conway, M.A., Cohen, G., and Stanhope, N.M. On the very long-term retention of knowledge. In *Journal of Experimental Psychology: General*, **120**, 395–409. (1991).

26. Degen, L., Mander, R., Salomon, G. Working with audio: integrating personal tape recorders and desktop computers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 413–418. (1992).

27. Dimitrova, N. Multimedia Content Analysis and Indexing for Filtering and Retrieval Applications. In Informing Science, Special Issue on Multimedia Informing Technologies. **2**(4). (1999).

28. Dumais, S.T., Cutrell, E., Cadiz, J.J., Jancke, G., Sarin, R. and Robbins, D.C. Stuff I've Seen: A system for personal information retrieval and re-use. *Proc. SIGIR 2003*. (2003).

29. Eagle, N., Singh, P., and Pentland, A. Common Sense Conversations: Understanding Casual Conversation using a Common Sense Database. Proceedings of Artificial Intelligence, Information Access, and Mobile Computing Workshop at the 18th International Joint Conference on Artificial Intelligence (IJCAI). Acapulco, Mexico. (August 2003).

30. Ebbinghaus, H. *Memory: A Contribution to Experimental Psychology.*, trans. by H.A. Ruber and C.E. Bussenius. 1964. New York: Dover. Original work published 1885.

31. Eldridge, M., Barnard, P., Bekerian, D. Autobiographical Memory and Daily Schemas at Work. *Memory*. **2**(1), 51–74. (1994).

32. Eldridge M., Sellen A., and Bekerian D., Memory Problems at Work: Their Range, Frequency, and Severity. Technical Report EPC-1992-129. Rank Xerox Research Centre. (1992).

33. Emnett, K. and Schmandt, C. Synthetic News Radio. *IBM Systems Journal*, **39**(3,4), 646–659. (2000).

34. Engen, T. and Ross, M.B. Lont-term memory of odors with and without verbal descriptions. *Journal of Experimental Psychology*. 100, 221–227. (1973).

35. Europen Union: Directive 2002/58/EC of the Eurpean Parliament and the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

36. Familiar Linux, http://familiar.handhelds.org/

37. Fernandez, R. *A Computational Model for the Automatic Recognition of Affect in Speech*. Ph.D. thesis. MIT program in Media Arts and Sciences. (February 2004).

38. Fertig, S., Freeman, E., and Gelernter, D. Lifestreams: an alternative to the desktop metaphor. *Proc. CHI '96*, 410–411. (1996).

39. Finkenauer, C., Luminet, O., Gisle, L., El-Ahmadi, A., Van Der Linden, M., and Philippot, P. Flashbulb memories and the underlying mechanisms of their formation: Towards an emotional-integrative model. In *Memory & Cognition*. **26**(3), 516–531. (1998).

40. Foulke, W., and Sticht, T.G. Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin*, **72**, 50–62, (1969).

41. Garofolo, J., Auzanne, C., and Voorhees, E. The TREC Spoken Document Retrieval Track: A Success Story. In *Proc. TREC 8*. 107–130. (1999).

42. Gemmell, J., Bell, G., Lueder, R., Drucker, S., and Wong, C., MyLifeBits: Fulfilling the Memex Vision, Proc. *ACM Multimedia '02*, Juan-les-Pins, France, 235–238. (2002).

43. Gerasimov, V. *Every Sign of Life*. Ph.D. thesis. MIT program in Media Arts and Sciences. (2003).

44. Gooden, D. and Baddeley, A.D. When does context influence recognition memory? in *British Journal of Psychology*, **71**, 99–104. (1980).

45. Gross, D. How many mobile phones does the world need? Slate. (June 2, 2004). http://slate.msn.com/id/2101625/

46. Hadis, M., *From generation to generation: family stories, computers and genealogy.* Masters Thesis. MIT Media Arts and Sciences. (2002).

47. Hayes, G.R., Patel, S.N., Truong, K.N., Iachello, G., Kientz, J.A., Farmer, R., Abowd, G.D. The Personal Audio Loop: Designing a Ubiquitous Audio-Based Memory Aid. To appear in the *Proceedings of Mobile HCI 2004: The 6th International Conference on Human Computer Interaction with Mobile Devices and Services.* (September 13–16, Glasgow, Scotland), 2004.

48. Healey, J. and Picard, R. StartleCam: A Cybernetic Wearable Camera. In *Proceedings of the Second International Conference on Wearable Computers.* (October 1998).

49. Heeman, P.A., Byron, D., and Allen, J.F., Identifying Discourse Markers in Spoken Dialog. *Proc. AAAI 1998 Spring Symposium on Applying Machine Learning to Discourse Processing.* (1998).

50. Henja, D. and Musicus, B.R., The SOLAFS Time-Scale Modification Algorithm, BBN Technical Report, (July 1991).

51. Hidden Cameras, Hidden Microphones: At the Crossroads of Journalism, Ethics, and the Law. http://www.rtnda.org/resources/hiddencamera/allstates.html

52. Hindus, D. and Schmandt, C. Ubiquitous Audio: Capturing Spontaneous Collaboration. *Proc. CSCW '92.* 210–217 (1992).

53. Johnson, C. Privacy Lost: The Supreme Court's Failure to Secure Privacy in That Which is Most Private – Personal Diaries. 33 *McGeorge L. Rev.* 129. (2001).

54. Kaye, J.N. *Symbolic Olfactory Display.* Masters Thesis, MIT Media Arts and Sciences. (May 2001).

55. Kletz, T. Lessons from Disaster: How Organizations Have No Memory and Accidents Recur. Institution of Chemical Engineers. Rugby, Warwickshire, UK. (1993).

56. Lame, http://lame.sourceforge.net/

57. Lamming, M., Brown, P., Carter, P., Eldridge, M., Flynn, M., Louie, P., Robinson, and P., Sellen, A. The Design of a Human Memory Prosthesis. *The Computer Journal.* **37**(3), 153–63 (1994).

58. Lamming, M. and Flynn, M. "Forget-me-not" - Intimate Computing in Support of Human Memory. In *Proceedings of FRIEND21, International Symposium on Next Generation Human Interface*, Megufo Gajoen, Japan (1994).

59. Lansdale, M., Edmonds, E. Using memory for events in the design of personal filing systems. in *International Journal of Man-Machine Studies.* **36**, 97–126. (1992).

60. "Larry King Live" Transcript (CNN), http://www.cnn.com/ALLPOLITICS/stories/1999/02/16/tripp.01/ (February 16, 1999).

61. "Larry King Live" Transcript (CNN), http://www.cnn.com/2001/ALLPOLITICS/02/09/tripp.lkl/ (February 9, 2001).

62. LeDoux, J.E. Emotion, Memory, and the Brain. In *Scientific American.* 50–57. (June 1994).

63. Levin, T.Y., Frohne, U., and Weibel, P. (eds.) *CTRL[SPACE]: Rhetorics of Surveillance from Bentham to Big Brother.* (MIT Press, 2002).

64. LifeBlog, Nokia, http://www.nokia.com/nokia/0,,54628,00.html

65. LifeLog, http://www.darpa.mil/ipto/programs/lifelog/

66. Lin, W. and Hauptmann, A. A Wearable Digital Library of Personal Conversations. *JCDL 2002*: 277–278 (2002).

67. Linton, M. "Memory for real-world events." In Norman, D.A. and Rumelhart, D.E. (eds.), *Explorations in cognition* (Chapter 14). San Francisco: Freeman. (1975).

68. Linton, M. Transformations of memory in everyday life. In *Memory Observed: Remembering in Natural Contexts*. U. Neisser, Ed. San Francisco. W.H. Freeman. (1982).

69. Liu F., Stern R., Huang X., and Acero A. Efficient Cepstral Normalization for Robust Speech Recognition. *Proceedings of ARPA Human Language Technology Workshop*, (March 1993).

70. Loftus, E.F. *Eyewitness Testimony*. Harvard Univ. Press, Cambridge, Massachusetts, (1996).

71. Loftus, E.F. and Marburger, W. Since the eruption of Mt. St. Helens, did anyone beat you up? Improving the accuracy of retrospective reports with landmark events. *Memory and Cognition*, **11**, 114–120. (1983).

72. Lucene, http://jakarta.apache.org/lucene/

73. Mani, I., Maybury, M.T. (eds.) Advances in Automatic Text Summarization. MIT Press. Cambridge, MA. (1999).

74. Mann, S. Wearable Tetherless Computer-Mediated Reality: WearCam as a wearable face-recognizer, and other applications for the disabled. In *Proceedings of AAAI Fall Symposium on Developing Assistive Technology for People with Disabilities*. (11 November 1996).

75. Marshall, A. Signal-Based Location Sensing using 802.11b. Advanced Undergradute Project. MIT Media Lab. (March 2003).

76. Monty, M.L. *Issues for Supporting Notetaking and Note Using in the Computer Environment*, Ph.D. thesis, Department of Psychology, University of Califomia, San Diego, CA (1990).

77. Moran, T.P., Palen, L., Harrison, S., Chiu, P., Kimber, D., Minneman, S., van Melle, W., and Zellweger, P. "I'll get that off the audio": A case study of salvaging multimediameeting records. *Proc. of CHI'97*. (1997).

78. Morgan, N., Baron, D., Edwards, J., Ellis, D., Gelbart D., Janin, A., Pfau, T., Shriberg, E., and Stolcke, A. The Meeting Project at ICSI. *Human Language Technologies Conference*. (2001).

79. Ng, C., Wilkinson, R., and Zobel, J. Experiments in Spoken Document Retrieval Phoneme N-grams. *Speech Communication, special issue on Accessing Information in Spoken Audio*, **32**(1-2), 61–77. (September 2000).

80. Ng K., Zue V.W. Phonetic Recognition For Spoken Document Retrieval. In *Proc. ICASSP*. 325–328. (1998)

81. Open Palmtop Integrated Environment (OPIE), http://opie.handhelds.org/

82. Orr, D.B., Friedman, H.L., and Williams, J.C., Trainability of Listening comprehension of Speeded Discourse." *Journal of Educational Psychology*, **56**, 148–156 (1965).

83. Orwant, J. Heterogenous learning in the doppelgänger user modeling system. *User Modeling and User-Adapted Interaction*, **4**(2), 107–130, (1995).

84. Rabiner, L., Juang, B., *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ. Prentice-Hall. (1993).

85. Reder, S., Schwab, R.G., The temporal structure of cooperative activity. In Proc. CSCW 1990. 303–316. (1990).

86. Rhodes, B. *Just-In-Time Information Retrieval*. Ph.D. Dissertation, MIT Media Lab (May 2000).

87. Ringel, M., Cutrell, E., Dumais, S.T., and Horvitz, E. Milestones in time: The value of landmarks in retrieving information from personal stores. *Proc. Interact 2003.* (2003).

88. Rogers, M.G., Bodily Intrusion in Search of Evidence: A Study in Fourth Amendment Decisionmaking. 62 *Ind. L.J.* 1181. (1987).

89. Rose, D.E. Mander, R., Oren, T., Poncéleon, D.B., Salomon, G., and Wong, Y.Y., Content awareness in a file system interface: implementing the pile metaphor for organizing information. In *Proceedings of the sixteenth annual international ACM SIGIR conference on Research and Development in Information Retrieval.* 260–269. (1993).

90. Sabbag, R. The Reluctant Retiree. *Boston Globe Magazine.* (November 2, 2003)

91. Salton, G., *Automatic Text Processing: the transfomation, analysis, and retrieval of information by computer.* Addison-Wesley (1989).

92. Schacter, D.L. The Seven Sins of Memory: Insights from Psychology and Cognitive Neuroscience. *American Psychologist.* **54**(3), 182–203 (1999).

93. Scharfe, E. and Bartholomew, K. Do you remember? Recollections of adult attachment patterns. Personal Relationships. **5**, 219–234. (1998).

94. Schmandt, C. The Intelligent Ear: An Interface to Digital Audio. *Proc. IEEE International on Cybernetics and Society,* IEEE, Atlanta, GA (1981).

95. Schmandt, C. *Voice Communication with Computers: Conversational Systems.* Van Nostrand Reinhold, New York. (1994).

96. Schmidt, R.A. and Lee T.D. *Motor Control and Learning. A Behavioral Emphasis.* 3rd edition. Champaign, IL: Human Kinetics. (1999).

97. Schwabe, G. Providing for Organizational Memory in Computer Supported Meetings. *Journal of Organizational Computing and Electronic Commerce,* 1999.

98. Shipman, F.M., Marshall C.C., and Moran, T.P., Finding and using implicit structure in human-organized spatial layouts of information. Conference proceedings on Human factors in computing systems. 346–353. (1995).

99. Siegler M.A., Witbrock, M.J., Slattery, S.T., Seymore, K., Jones, R.E., and Hauptmann, A.G. Experiments in Spoken Document Retrieval at CMU. *Proc. TREC 6.* (1997).

100. Singh, J., Peck, R.A. and Loeb, G.E. Development of BION Technology for functional electrical stimulation: Hermetic Packaging. In *Proc. IEEE-EMBS* (Istanbul, Turkey), (2001).

101. Sinoff, G. and Werner, P. Anxiety disorder and accompanying subjective memory loss in the elderly as a predictor of future cognitive decline. *Int J Geriatr Psychiatry.* **18**(10), 951–9 (Oct 2003).

102. Smeaton, A.F., Over, P. The TREC-2002 Video Track Report. *Proc. TREC 11.* (November 2002).

103. Spence R. Rapid, serial and visual: a presentation technique with potential. *Information Visualization,* **1**(1), 13–19. (2002).

104. Stark, L., Whittaker, S., and Hirschberg, J. ASR satisficing: the effects of ASR accuracy on speech retrieval. *Proc. International Conference on Spoken Language Processing.* (2000).

105. Sticht, T.G. Comprehension of repeated time-compression recordings. *The Journal of Experimental Education,* **37**(4), (Summer 1969)

106. Stifelman, L., Arons, B., and Schmandt, C. The audio notebook: paper and pen interaction with structured speech. In *Proceedings of the SIG-CHI on Human factors in computing systems.* 182–189. (2001).

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

108. Tulving, E. The effects of presentation and recall on material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, **6**, 175–184. (1967).

109. Tulving, E. and Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, **5**, 381–391. (1966).

110. United Stated Electronic Communications Privacy Act of 1986, 18 U.S.C. 2510.

111. Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT ACT) Act of 2001. (Public Law 107-56; 115 Stat. 272)

112. Vemuri, S., DeCamp, P., Bender, W., Schmandt, C. Improving Speech Playback using Time-Compression and Speech Recognition. In *Proc. CHI 2004*.

113. Vemuri, S., Schmandt, C., Bender, W. An Audio-Based Personal Memory Aid. To appear in Proc. *Ubicomp 2004*.

114. ViaVoice, http://www-3.ibm.com/software/speech/

115. ViaVoice, Frequently Asked Questions http://www.wizzardsoftware.com/voice/voicetools/dictationforsdkfaq.htm#What%20is%20Phrase%20Score.

116. Wagenaar, W.A. My Memory: A study of Autobiographical Memory over Six Years. In *Cognitive Psychology*. **18**, 225–52 (1986).

117. Wactlar H.D., Hauptman A.G., and Witbrock M.J., Informedia News-On Demand: Using Speech Recognition to Create a Digital Video Library. Tech Report. CMU-CS-98-109 (March 1998).

118. Warrington, E.K. and Sanders, H.I., The fate of old memories. In *Quarterly Journal of Experimental Psychology*, **23**, 432–42. (1971).

119. Wechsler, M., Munteanu, E., Schäuble, P.: New Techniques for Open-Vocabulary Spoken Document Retrieval. *Proc. SIGIR 1998*. 20–27. (1998).

120. Weiser, M., The Computer for the Twenty-First Century. *Scientific American*, 94–10 (September 1991).

121. Whittaker, S. and Amento, B. Semantic Speech Editing. In *Proc. CHI 2004*. (2004).

122. Whittaker, S., Hirschberg, J., Amento, B., Stark, L., Bacchiani, M., Isenhour, P., Stead, L., Zamchick G., and Rosenberg, A. SCANMail: a voicemail interface that makes speech browsable, readable and searchable. *Proc. CHI 2002*, 275–82 (2002).

123. Whittaker, S., Hirschberg, J., Choi, J., Hindle, D., Pereira, F., and Singhal, A. SCAN: designing and evaluating user interfaces to support retrieval from speech archives. *Proc. SIGIR99*. 26–33. (1999).

124. Whittaker, S., Hyland, P., and Wiley, M. Filochat: handwritten notes provide access to recorded conversations. In *Conference proceedings on Human factors in computing systems*. 271–277. (1994).

125. Whittaker, S. and Sidner, C. Email overload: exploring personal information management of email. In *Proceedings of CHI'96 Conference on Computer Human Interaction*. 276–283. (1996).

126. Wilding, E.L., Rugg, M.D. An event-related potential study of memory for words spoken aloud or heard. Neuropsychologia. 35(9), 1185–95 (1997).

127. Williams, L. SenseCam. http://research.microsoft.com/research/hwsystems/

128. Witbrock, M., http://infonortics.com/searchengines/boston1999/witbrock/index.htm, Lycos (1999). [link valid as of Jan 1, 2004]

129. Wong, B.A., Starner, T.E., and McGuire, R.M. Towards Conversational Speech Recognition for a Wearable Computer Based Appointment Scheduling Agent. GVU Tech Report GIT-GVU-02-17. (2002).