Towards Conversational Speech Recognition for Scheduling Agents on Wearable Computers

Cornelis M. Snoeck, Thad E. Starner, Benjamin A. Wong, R. Martin McGuire, and Jack Greundler
College of Computing and GVU Center
Georgia Institute of Technology, Atlanta, GA, USA 30332-0280

Mobile speech interfaces that are socially acceptable and technically feasible are difficult to design. This paper presents a first step in exploring conversational interfaces that use socially applicable speech to cue an appointment scheduling agent on a wearable computing platform. We present a user study of current appointment scheduling mechanisms which reveals a relationship between the time required to access a device and the probability that the user will choose an alternative method of recording the appointment. In addition, the study demonstrates that many subjects postpone entering appointments into their calendar until after the end of the conversation. These results inform the design of two prototype systems: the Calendar Navigator Agent (CNA) and Dialog Tabs.

The CNA monitors the user's speech and displays calendar information on a head-up display as the conversation progresses. We show that such an interface has the potential to speed the appointment scheduling process significantly. Dialog Tabs attempts to minimize user distraction by buffering and indexing appointment conversations for later processing.

Categories and Subject Descriptors: []:

General Terms:
Additional Key Words and Phrases: appointment scheduling, intelligent agents, context awareness, speech recognition, wearable computing

1. INTRODUCTION

Unlike PDAs and laptops which reside in pockets or briefcases, wearable computers enable quick interaction through mobile keyboards and high-resolution head-mounted displays (see Figures 1 and 2). By distributing sensors on the user's body (e.g. a cameras, microphones, accelerometers, etc.), a wearable can continually observe the environment from the same perspective as the user [Starner 1999]. Such access to the user's context enables the creation of pro-active wearable agents. Here, we wish to utilize audio input to create an agent that assists the user with performing daily activities by listening to and acting on the user's conversations.

Speech is ubiquitous in today's office environment; office workers spend 35-80%

This work funded in part by Starner's NSF Career Grant #0093291 and the NIDRR Wireless RERC.

of their time in spoken conversation. High-end managers are generally at the top end of this range. In addition, opportunistic communication may account for up to 93% of these managers' work days [Whittaker et al. 1994]. Currently, much of the information from these conversations is lost, but prototype systems designed for capturing and accessing verbal communication in stationary environments have been shown to utilize speech better [Abowd et al. 1998]. In a mobile environment, unfortunately, automatic speech recognition is difficult due to variable background noise and adaptive user behavior. For example, when a person speaks in a noisy environment, he changes his articulation and volume to compensate for the background noise [Junqua 1993]. The user's change in speech decreases recognition accuracy since most speech recognizers are not trained for such conditions. However, even when a flawless transcription of the user's speech can be obtained, complex reasoning may still be required to extract useful information from the conversation [Schmandt 1994b].

Recently, popular press articles and conference panel sessions have been critical of speech systems [Newman 2000; James 2002]. Such articles may be in response to consumer disappointment in commercial dictation systems as well as a reaction to earlier concept videos that portrayed anthropomorphic agents addressed through speech. However, conversational system researchers have written articles about the limitations of these systems and where they are most useful for many years [Schmandt 1994a; Karat et al. 1999; Yankelovich et al. 1995; Oviatt 1999; Danis et al. 1994]. Shneiderman provides a brief overview of the issues in his 'Limits of Speech Recognition" [Shneiderman 2000]. Cohen and Oviatt provide a list of situations when speech may be advantageous [Cohen and Oviatt 1995]:

(1) When the user's hands or eyes are busy.
(2) When only a limited keyboard and/or screen is available.
(3) When the user is disabled.
(4) When pronunciation is the subject matter of computer use.
(5) When natural language interaction is preferred.

In the past, wearable computer research has concentrated on the first three situations [Najjar et al. 1997; Smailagic and Siewiorek 1994; Ockerman et al. 1997; Stein et al. 1998; Starner et al. 1998; Collins et al. 1977; Ross and Blasch 2000; Upton 1968]. However, we wish to begin to explore the last situation — the use of natural language interaction. Related projects have avoided speech recognition and stored the audio directly, using other cues, such as pen strokes, location, or time of day, for indexing the audio [Stifelman et al. 1993; Stifelman 1996; Whittaker et al. 1994; Wilcox et al. 1997; Abowd et al. 1998]. Roy et al. provide an overview of these methods [Roy et al. 1997]. Such systems are designed for situations in which the amount of spoken information is overwhelming (e.g. attending a conference) and speech transcription is not feasible due to the complexity of the utterances. However, we have observed a domain in which tasks adhere to a much more limited script; the process of appointment scheduling. In addition, the large markets for paper-based day planners and PDAs indicate a desire for assistance in appointment scheduling. Thus, we focus our efforts on calendar agents.

Fig. 1. Wearable computer user with MicroOptical SV-3 640x480 color display, 800MHz CharmIT Pro, and Twiddler 2 one-handed chording keyboard.

Fig. 2. Prescription eyeglasses with 320x240 monochrome display integrated in the user's left lens (1997 prototype).

1.1 Paper Outline

In the first section, the results of a user study at the Georgia Tech Student Center are presented which uncover commonly used scheduling strategies. Remarkably, a large number of participants was found not to use the device they claimed to use beforehand. The reasons of this behavior were explored to inform the design of two scheduling agents.

The first prototype described in Section 3, called Calendar Navigator Agent (CNA), tries to use the quick accessibility of the wearable computer to minimize the time required to enter an appointment. It automatically navigates the user's calendar based on the natural person-to-person speech used to negotiate an appointment. Thus, the user does not have to interrupt the conversation to access the CNA. In addition to the expected problems due to speech recognition errors, the CNA raises other challenges which are discussed in Section 4. Therefore, a second prototype, Dialog Tabs, was developed to explore these issues further.

Dialog Tabs, described in Section 5, uses another commonly observed scheduling strategy. Instead of updating the user's calendar during the conversation, Dialog Tabs mimics the use of memory and scrap paper by capturing audio for later processing by the user. This technique combines the speed of access of an unstructured scheduling mechanism such as scrap paper and the convenience of an electronic device that enables rapid search without accidentally losing or forgetting appointments. Finally, Section 6 suggests future work in exploring these calendar agents.

2. CALENDAR USER STUDY

To gain insight into currently used scheduling strategies, a user study was performed inside the main entrance of Georgia Tech's Student Center, asking passersby to volunteer as subjects.

2.1 Study Setup

The study consisted of two parts: a short questionnaire and an appointment scheduling session. The questionnaire requested demographic information, an exhaustive list of all calendar systems used by the subject, the primary calendar system the subject uses when away from his desk, how long the subject has been using this calendar system, and how many appointments the subject schedules per week. Eight Likert scale questions (ranging from 1 for strong agreement to 7 for strong disagreement) were used to elicit the subjects' opinions on their calendar systems.

Each participant was asked to sit at a table for an interview with one of our researchers to perform timing tests on appointment scheduling practices. While it was intended for the subjects to perform the appointment scheduling tasks after answering the questionnaire, a small number of subjects answered the questionnaire after completing the tasks. A single researcher performed all of the appointment scheduling with the subjects, using a script of four tasks:

A. Schedule an appointment for a date seven days in the future.

B. Schedule an appointment for a date three months in the future.

C. Schedule an appointment for tomorrow.

D. Reschedule the second appointment to the next day.

Subjects were asked to re-stow their preferred scheduling device if it was retrieved in anticipation of the scheduling tasks. However, once the subject had retrieved his device during the interview, he was not told to re-stow it before subsequent tasks. Subjects were encouraged to schedule the appointments as if they were of significant importance and scheduling conflicts were resolved as part of the task but not included in the recorded times.

In order to capture timing data accurately, the experiment was videotaped with two cameras: one pointing forward toward the test subject and a second pointing down at where test subject was likely to place their scheduling device while in use (Figure 3). The cameras were time synchronized, and both recorded audio.

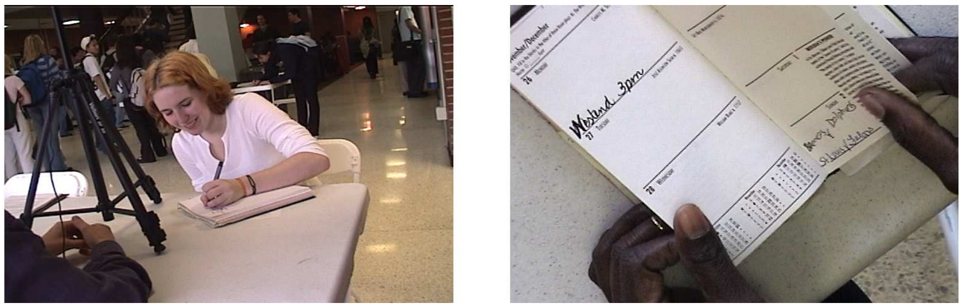

Fig. 3. Typical images from the forward and downward pointing video camera taken during appointment tasks.

Although it is difficult to extract data from video, the alternative of instrumenting each individual's scheduling device to record data would have sacrificed ecological validity. Additionally, this practice would not have allowed the observation of any discrepancy between the claimed and actual scheduling device used.

Based on an extensive preliminary examination of the data, a protocol was formulated to extract the time required to interact with each calendar mechanism. The overall time was divided into three parts: physically retrieving the device (e.g. removing the device from the user's pocket), navigating the device's interface (e.g. selecting the correct date and time), and entering the appointment (e.g. transcribing participant and location information). A single researcher used the protocol to transcribe all scheduling tasks on the recorded video. Samples were checked for accuracy by three additional researchers.

2.2 Study Results

A total of 138 subjects participated in the study, with a predominance of young male students (88% age 18-25, 70% male, 90% students). Most of the subjects used a paper-based planner, memory, scrap paper, or a PDA, and therefore the result analysis will focus on these calendar mechanisms. Timing results will only be considered from the first scheduling task because, by the second task, the participants may have already retrieved their calendar device.

2.2.1 *Device Disuse.*

A significant fraction of subjects does not use the device they claim as primary when scheduling appointments. Figure 4 shows the distribution of the devices that subjects claimed to use on the questionnaire and devices that were used during the completion of the first task. Note the differences between the graphs for planners, memory, and scrap paper. The following subsections will explore the study results by explaining the reasons for this observation of device disuse.

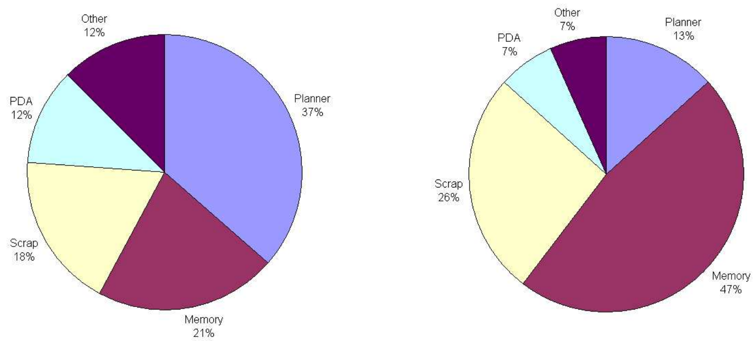

Fig. 4. Left: devices claimed as primary on the questionnaire. Right: devices actually used during the appointment tasks.

2.2.1.1 *Disuse vs. Retrieval and Navigation Time.* Table I illustrates the mean retrieval, navigation, entry, and total scheduling times of various calendar mechanisms recorded during the subjects' appointment scheduling tasks. Due to its nature, interaction times for memory are difficult to perceive. Therefore, for the purposes of this study, we assume interaction times for memory are near zero compared to the other mechanisms.

Note that the total scheduling time is not significantly different between the different mechanisms. In contrast, the sum of retrieval and navigation times shows significant variance between devices. WE HYPOTHESIZE THAT THE TIME REQUIRED TO RETRIEVE THE DEVICE PHYSICALLY PLUS THE TIME TO NAVIGATE THE DEVICE'S INTERFACE IS RELATED TO ITS RATE OF DISUSE.

Device	*Retrieval*	*Navigation*	**Retr. + Nav.**	*Entry*	*Total*
Scrap	17.8		**17.8**	18.1	35.9
Planner	11.8	7.6	**19.4**	12.5	31.9
PDA	11.0	12.7	**23.7**	14.0	37.7

Table I. Timing results per commonly used device (average in seconds). Navigation time for scrap paper is considered effectively zero.

Table II shows the claimed versus actual device usage in more detail. For example, the second column of the first row shows that nine subjects who claimed to use scrap paper on the questionnaire used memory during the first task of the interview instead. The table rows are sorted by the average amount of time required to retrieve the device and navigate its interface. Note that very few entries cross the diagonal boundary in the table, indicating that almost all users who did not use their claimed mechanism switched to a faster device in practice. Figure 5 provides a visual rendering of this phenomenon for the main mechanisms used in the study. Devices with lower times tend to have higher actual to claimed usage ratios. We conclude that retrieval and navigation time are related to the use of a scheduling device.

The more structured scheduling mechanisms, PDAs and planners, require the most time to reach the appointment entry location (retrieval + navigation time).

	Claimed Usage: Memory	Scrap	Planner	PDA	Other	# Used
Actual Usage: Memory	24	9	16	4	4	57
Scrap	1	13	13	1	4	32
Planner			14	1	1	16
PDA				8		8
Other	1		1		6	8
# Claimed	26	22	44	14	15	121

Table II. Claimed vs. Actual Device Usage on the first task. Rows are sorted in order of the average amount of time required to retrieve the device and navigate its interface. Zeros have been left blank to better show the pattern of device abandonment from slower to faster devices.

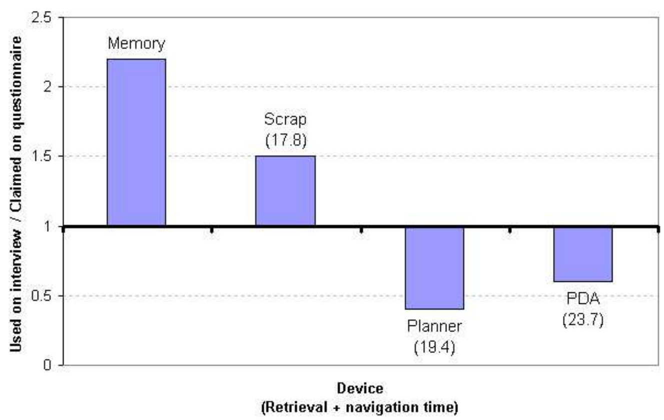

Fig. 5. Actual vs. claimed usage ratio of devices (ordered by retrieval plus navigation time).

In contrast, scrap paper requires significantly more time for appointment entry, probably due to a need to note the time and day information that is specified through the navigation step for PDAs and planners.

Scheduling devices are often manipulated while the user is gathering appointment information through speech. For example, a subject may flip through the pages of his paper calendar during a conversation. Therefore the times measured in this study are probably higher than those required for each device when the user does not have to divide his attention. Also, device retrieval times might vary greatly, depending on where the users stored their device before participating in the study (e.g. backpack, pocket, etc.). The study was designed to preserve these effects since it models the actual practices of users.

2.2.1.2 *Disuse vs. Cognitive Load.* Interacting with a scheduling device during a conversation requires the user to divide his attention between the interlocutor and the device, which may cause an increase in cognitive load . Various researchers have observed related cognitive load effects in the literature [Shneiderman 2000; Karl et al. 1993; Schacter 2001; Wickens 1984; Blackwood 1997]. We hypothesize that devices that increase peak cognitive load are not preferred during a scheduling conversation. Some of the study data and anecdotal evidence support this hypothesis.

For the first task, less experienced day planner users were found to be more likely to abandon their interface than those with more experience ($\mu_{disuse} = 27.0$; $\mu_{use} = 49.1$; p = 0.012) [1] This effect may be the result of an attempt to avoid the cognitive load of navigating a less familiar device during the conversation. The abandonment of a PDA is not significantly correlated with experience. However, many PDA users have been observed to make filler conversation while accessing their devices (e.g. "Let me get that down...November 3rd...there it is...OK...what room again?"). Perhaps the user does not want to slow down or interrupt the conversation but is forced to devote most of his attention to navigating his calendar device.

2.2.1.3 *Disuse vs. Forgetting the Device.* Some PDA and day planner users explain their disuse behavior by indicating that they left their device at some other location. Figure 6 shows the percentage of subjects who did not use their claimed primary mechanism during the first and second tasks. A high and consistent rate of disuse by planner users was observed for all four tasks which supports this explanation. Yet users forgetting their devices can not explain the difference in disuse rates between the tasks for PDAs and scrap paper. This figure demonstrates that a number of PDA and scrap paper users had their device with them, yet retrieved the device for the second task and not for the first.

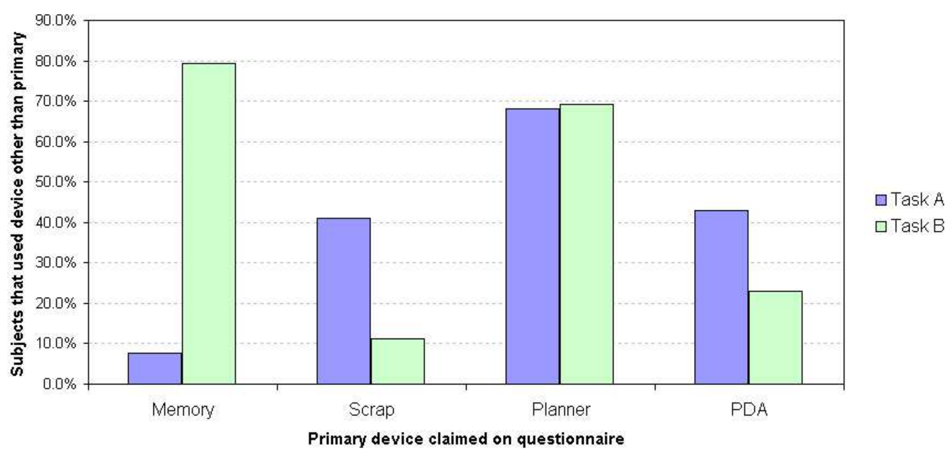

Fig. 6. Abandonment rates of claimed primary device during the first (next week) and second (in three months) scheduling tasks.

Perhaps when these users realized after the first task that the interviewer was going to ask them to schedule several appointments, the benefit of using their PDA or scrap paper became apparent. In other words, the barrier of access time or cognitive load might have been lessened by the batch nature of the tasks. A more likely explanation is that these subjects were willing to rely on their memory

[1] In general, we use the Welch modification to the t–test to derive our p–values.

instead of their device to schedule an appointment in the near future (task A), but needed the device to schedule an appointment several months in the future (task B). Corroborating, every subject who switched to their primary device on the second task used memory to complete the first task. This hypothesis is also supported by the large percentage of claimed memory users who switched to more permanent memory mechanisms (scrap paper, planner, phones, writing on skin, etc.) for the appointment made in the distant future. The use of a temporary mechanism to "buffer" appointments is examined in more detail in the next section.

2.2.2 *Buffering of Appointments.* In the previous section it was shown that a large number of subjects did not use the device claimed as primary on the questionnaire. Of those subjects, 93% used memory or scrap paper during the interview.

In general, two scheduling strategies were observed for subjects that did not use their primary scheduling device:

(1) Buffering the obtained appointment information on a temporary device for later reconciliation with the primary device.

(2) Using a different device in parallel with the primary scheduling device.

Cued by results from a pilot study, we asked memory and scrap paper users whether they would transfer the appointment information they just received to another mechanism (e.g. planner, PDA, etc.); approximately 42% of the subjects admitted to using the this practice for the first task. With the second task, which required confirming an appointment further in the future, 86% of the users said they would transfer the information to another device. Intuitively, people may use temporary devices such as scrap paper and memory concurrently with their primary scheduling system for short term scheduling (within approximately a week), whereas longer term planning requires more strict reconciliation.

We observe that PEOPLE POSTPONE THE ENTRY OF APPOINTMENTS DURING A CONVERSATION BY USING MEMORY OR SCRAP PAPER AS TEMPORARY DEVICES. We hypothesize that the goal of this postponement behavior is to prevent disruption of the conversation. Section 2.2.1 indicates that these disruptions may be caused by large retrieval and navigation times or increased cognitive load.

2.2.3 *General Study Observations.* The answers to the Likert scale questions were only surprising in that they were generally consistent across the various devices. No matter which system was used, subjects were inclined to indicate that their mechanism was appropriate, sufficient, and somewhat necessary for reminding them of appointments. While the questions related to mechanism effectiveness had slightly positive scores, questions related to ease of use and speed of access were strongly positive with little variance. Yet, the timing observations made in the mock scheduling tasks suggest that significant improvements can be made.

The notable exception to the consistency of the Likert answers was that PDA users overwhelmingly rated their system as expensive (p ¡ 10^{-5} compared with the users of other devices). This observation may explain why PDA users were less likely to abandon their device even though it required more time to retrieve and navigate than planners. Perhaps PDA users felt that they had invested a significant amount of money in their device and were reluctant to abandon them.

Device	*# appointments*
PDA	9.6
Planner	8.7
Memory	6.7
Scrap	5.4

Table III. Average number of appointments per week per device.

As reported in Section 2.2.1 many of our PDA and planner subjects did not use their claimed device during the appointment tasks. On the questionnaire many more subjects reported having used PDAs or planners in the past but did not consider these devices their primary calendar mechanism while they were away from their desks. Thus, these subjects, for some reason, decided not to use PDAs or planners. The numbers are significant, representing 36% more potential PDA users and 45% more potential planner users that would be added to the claimed usage table above if they had declared these devices as their primary mechanism. Of these potential users, at least 72% reported less informal mechanisms (e.g. scrap paper, memory, etc.) as primary. This evidence of abandonment corroborates anecdotes from former PDA users who claim that the effort required to maintain the data in their devices outweighed its benefits.

Given the issues suggested in previous sections, why are PDAs and planners used? As shown in Table III, PDA and planner users in our study averaged more appointments per week than the other common mechanisms. Perhaps such users feel that they have too many, or too far removed, appointments to remember them. In our Likert results, memory users agreed more strongly than planner users ($\mu_{memory} = 4.12$; $\mu_{planner} = 5.05$; p = 0.026) that they still forget or are late to appointments more often than they would like. In our questionnaire, PDA users perceive that their system more reliably reminds them of appointments than subjects who use their own memory ($\mu_{PDA} = 2.43$; $\mu_{memory} = 3.35$; p = 0.049), and PDA users expressed that the active reminder functions of their PDAs were desirable.

2.2.4 *Further study.* This study has provided insight into how calendar mechanisms are used and users' opinions about these systems. However, the queried population is not appropriate for studying the effects of large numbers of appointments. We hope to improve our study methodology and attempt the study with a population, such as the business travelers at Atlanta's Hartsfield airport, that is more likely to purchase electronic calendar aids.

Even with the demographic bias of a technical university's student center population, we feel that we have gained a better understanding of the flaws of current mobile calendar systems and use this as guidance for our prototype designs.

3. SCHEDULING AGENTS

In the next sections, the user-centered design of two scheduling agents for wearable computers is presented, motivated by the results of the previously described calendar user study. As shown in Section 2.2.1.1, scheduling devices with low retrieval and navigation times are preferred during face-to-face conversation. Generally, appointments cannot be scheduled without a dialog between the participants making the appointment. Thus, the most lightweight and immediately accessible scheduling assistant would use this dialog as input. Without prompting from the user, the assistant would listen to the scheduling dialog and extract the information necessary to record the appointment. Such an agent would only affect the flow of conversation for conflicting meeting times or confirmation. Unfortunately, in order to operate, such a computational assistant would require speech recognition on unconstrained language. User modeling may also be necessary to understand the context of a given appointment. For example, if a user ordinarily meets with a colleague at 2PM but wants to reschedule, he might say "Can we meet at 3PM instead?" The proper behavior of the assistant would be to cancel the 2PM meeting; however, this information was not explicitly spoken at any time. Furthermore the assistant would require a common knowledge representation to determine an appointment time from a sentence like "Let's meet before Starbucks closes."

We suggest two speech-based calendar agents that are appropriate in the conversational context of creating appointments, but do not require user modeling or knowledge representation. Our work is influenced by the use of speech recognition in the SCANMail system [Whittaker et al. 2002] and by Olsen's use of speech interfaces to access large lists of data [Olsen and Peachey 2002].

With our agents, the user purposefully restricts his speech to a known set of terms and phrases for which the speech recognizer is trained. These phrases are chosen to cue the agent while simultaneously informing the user's conversational partner in a socially applicable manner. For example, the phrase "Can we meet June 20th?" performs a conversational function as well as cues the user's agent as to which day is being discussed. This "dual-use" speech might at first seem to encourage stilted conversation. However, we observed in the calendar user study that schedule negotiation often follows predictable scripts. Codifying these scripts into an appropriate vocabulary and grammar that cues agents is an interesting research opportunity. The result must reduce the perplexity of the speech recognition problem while maintaining the patterns of natural dialog. The described agents below illustrate these principles, but extensive work is necessary to obtain the level of interaction of similar test beds such as ATIS [Hemphill et al. 1990; Kubala et al. 1994], Jupiter [Zue et al. 2000], or HMIHY [Walker et al. 2000].

We also simplify the speech recognition problem by using push-to-talk techniques to allow the user to specify to which parts of the conversation the computer should attend. To accomplish this purpose a button is mounted on the user's clothing in an easily accessible location. In addition, we employ a wearable computer equipped with a head-up display (HUD) to provide rapid visual feedback to the user during the schedule negotiation process (Figure 2).

3.1 Calendar Navigator Agent Operation

The first prototype, the Calendar Navigator Agent (CNA), is a calendar application similar to Microsoft Outlook that is operated by speech, keyboard, or mouse and is displayed on the HUD during the social interaction. The phrases that can be uttered to control the calendar are constructed from a limited grammar that was designed to emulate the conversations observed in the calendar study.

The operation of the CNA is illustrated in more detail by an example dialog. In this particular dialog the interlocutor (I) initiates the scheduling conversation.

Bold text indicates speech uttered by the wearable computer user (W) while holding down the push-to-talk button:

I: Can we meet in the second week of February?

The wearable user now mounts his microphone and HUD (usually already present), and responds appropriately while pressing the push-to-talk button. The calendar application window automatically pops up.

W: **Any particular day of the second week of February?**

After the user releases the button the recognized sentence is displayed in the lower right corner of the HUD (Figure 7). This assists the user in estimating the accuracy of the speech recognizer in the current environment. Also note that the appropriate week is highlighted in the month overview and all scheduled appointments for that week are displayed automatically.

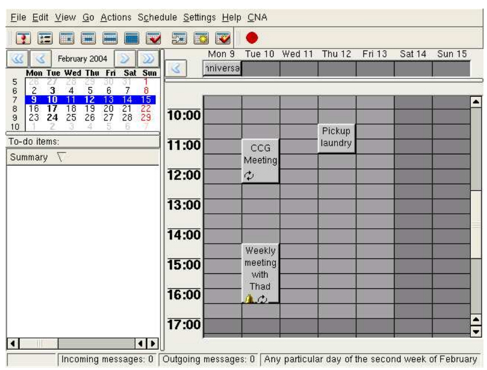

Fig. 7. Upon hearing the user say "Any particular day of the second week of February?" the CNA displays a summary of the second week of February on the user's head-up display.

I: How about Monday?

The user quickly perceives that he already has a whole-day appointment that Monday, but he would like to have a closer look at Tuesday.

W: I'm busy all day on Monday. **I could meet Tuesday though.**

Note the user only holds down the push-to-talk button during the second sentence. In response, the application zooms in on Tuesday, providing more detailed information on previously scheduled appointments (Figure 8).

I: Tuesday at one then?

Finally the user constructs a sentence with the appointment time in conjunction with the attendee's name:

W: **One o'clock it is, Peter.**

An appointment entry dialog box as in Figure 9 appears with the meeting time and summary fields already filled in, based on the user's utterance. If required,

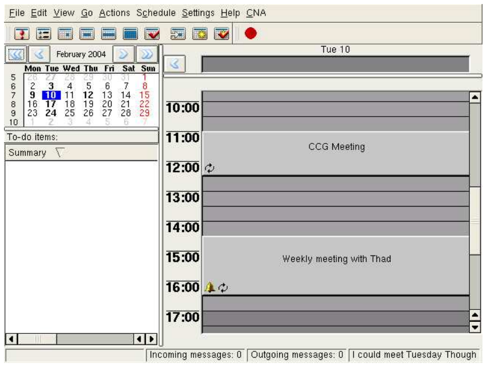

Fig. 8. The user saying "I could meet Tuesday though" results in the CNA displaying a summary for Tuesday.

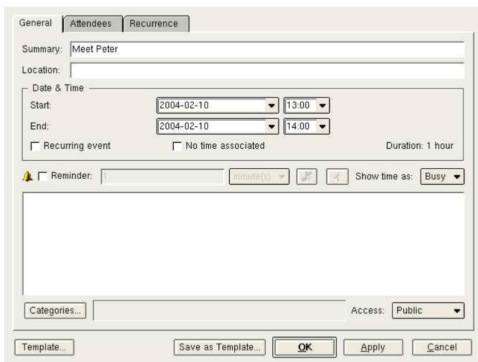

Fig. 9. When the user says "One o'clock it is, Peter" the CNA enters the appointment into the user's calendar.

additional information on the event can now be entered using the Twiddler, a one-handed chording keyboard. The user finishes the scheduling task by closing the dialog box manually or uttering the phrase:

W: **I'll see you later.**

Note that a limited grammar was sufficient to provide the CNA with appropriate information and was also used to resolve a scheduling conflict. The following example conversation shows that it is also possible to have the wearable computer user start the dialog:

W: **Could we meet next week on Thursday?**
I: Sure, what time?
W: **How about 11 o'clock?**
I: OK, see you then, Alice.
W: **See you then, Bob.**

The user can also navigate to an appointment date relative to the date that is currently being discussed:

W: **Could you meet tomorrow at eleven?**
I: That's not possible.
W: **Perhaps the next day?**
I: The day after tomorrow ... that's possible I guess, what time?
W: **Eleven o'clock?**
I: That's fine.
W: **I'll pencil you in, Sylvia.**

3.2 CNA Timing Test

In Section 2.2.1.1 we showed that high retrieval and navigation times are related to the disuse of devices such as PDAs and planners. To get an indication of the viability of the CNA, a controlled environment timing experiment was performed in which the speed of the CNA was compared with several scheduling devices that were observed during the calendar user study. This is a one-participant pilot study that minimizes the degrees of freedom by solely focusing on timing, ignoring other potential issues of the CNA which will be discussed in Section 4. The pilot study's primary goal was to indicate whether further pursuit in this line of work may be fruitful.

We compared an HP Jornada 586 Pocket PC with the Windows CE calendar and letter-by-letter handwriting recognition input; a CNA prototype without name recognition; and a paper-based calendar with one day per two pages, a tab to indicate the current day, and no index tags for days, weeks, or months. The wearable computer is assumed to be instantly accessible by its user. In other words, the head-up display is already mounted and the push-to-talk button is readily available as in shown in Figure 1.

To help respect naturalness in speech, we constructed transcriptions of common interactions from the video data of the first three tasks in the user study, i.e. scheduling an appointment for a date seven days in the future, three months in the future, and tomorrow. For each task a transcription with and without a scheduling conflict was constructed. Example dialogs were shown in the previous section. Each dialog was performed twice for each device, alternating the person that starts the conversation. This procedure resulted in a total of twelve dialogs that were re-enacted by two researchers and videotaped to enable exact timing data extraction.

The researcher that performed the timed operations (one of the authors) was trained to be an expert user for all three scheduling device types and re-performed trials with spelling, navigation or speech recognition mistakes[2]. Before every trial the devices were reset to the current day.

The timing data was extracted from the video with the same protocol used in the calendar user study. The average retrieval, navigation and entry times are shown in Table IV. The CNA has no retrieval time since the device is already present.

On average the PDA was 47% slower than the CNA ($\mu_{PDA} > \mu_{CNA}$; p = 0.000013) and the paper based planner was 20% slower than the CNA ($\mu_{planner} > \mu_{CNA}$; p = 0.080). Notice that the CNA appears to take longer to navigate despite

[2]Inaccurate entries were not immediately corrected, instead the experiment continued with other dialogs to prevent the expert user from remembering hand movements or screen positions as a result of successive runs of the same dialog on a single device.

Device	*Retrieval*	*Navigation*	*Entry*	*Total*
CNA		13.7	4.3	18.0
Planner	9.3	7.9	4.3	21.5
PDA	8.7	10.3	7.5	26.5

Table IV. Timing results for an expert using a PDA, planner, and CNA (average in seconds) for appointment scheduling based on re-enacted dialogs from the calendar study.

the fact that the navigation time only consists of the duration of the negotiation phase of the conversation. The cause of this phenomenon is that the PDA and planner were retrieved in parallel with the conversation, and therefore the negotiation phase is distributed among the retrieval and navigation times for these devices. The entry time for the version of the CNA used for this experiment includes the time it took to enter names manually with a Twiddler keyboard. Note that with appointment participant name recognition as illustrated in Section 3.1 entry times should significantly decrease.

Also note that four of the twelve tasks were scheduled for the next day, which may be an unreasonably large proportion compared to average calendar use. Due to the planner's index tab designed for fast access to the current day, our testing provided the paper based planner a distinct advantage in navigation time because of the physical proximity of the next day to the current day. When ignoring the next-day-appointment tasks, the paper-based calendar becomes 35% slower than the CNA ($\mu_{planner} > \mu_{CNA}$; p = 0.01734).

One might argue that the timings of an "expert" user in a laboratory study will vary significantly from what might be expected in practice. However, the timing results of PDAs and paper-based planners in this preliminary experiment are consistent with comparable trials in the calendar user study; on average the data is within one standard deviation of the user study data mean, except for PDA navigation times which are on average 1.3 standard deviations lower than the mean of the user study data. Even though this pilot experiment is limited to one expert user and focuses only on timing, the results show that the CNA has the potential to reach lower interaction times than the more conventional scheduling devices in the study.

4. CHALLENGES FOR "LISTENING IN" INTERFACES

Interfaces that attend and act upon conversational speech between the user and others in the environment raise legal, social, psychological and technical issues. In this section these concerns with respect to the Calendar Navigation Agent are addressed.

4.1 Privacy and Legal Concerns

In most areas of the United States, recording of conversations in electronic media is permissible if at least one person in the conversation is aware of the recording. Thus, there is no legal restriction to a wearable computer user knowingly recording audio in his environment in these areas. However, in twelve states all participants in a conversation must give consent to recording for most situations. Such legal restrictions would seem to prohibit the use of agents that record conversational audio unless the system made clear that the recording was happening (for example, the blinking red "recording" light on camcorders).

In addition, the convention for current conversational practices assumes that audio is not being recorded by the participants. According to [Strubb et al. 1998], the public is more concerned with the surreptitious recording of audio than of video, and recording of another person's speech without their consent may be considered a serious breach of etiquette.

The CNA has been consciously designed with regard to these issues of privacy. It uses a noise-canceling microphone which attenuates speech from other people to an essentially inaudible level. In informal experiences of using such speech systems on a daily basis in the Georgia Tech environment, we have found that other members of the community understand and accept this explanation readily. In addition, the CNA discards the audio after the voice recognition system has interpreted it. Since speech recognition systems can not recognize speech at low levels, the CNA automatically disregards others' speech in the environment. Furthermore, even if a very good signal was somehow obtained of others' speech in the environment, the limited vocabulary and grammar of the CNA would prevent the parsing of any meaningful phrases from the signal.

Another prototype that will be explored in Section 5, Dialog Tabs, does store audio. Again the "noise-canceling" microphone can be employed to avoid legal and social problems. By simply not recording audio that is below a certain energy threshold, the system avoids capturing other people's conversations while still allowing the system to work for the wearer. Others in the user's environment might still object to the presence of the microphone and the user's mere capability of recording audio. However, in practice, society has already accepted the ubiquitous microphones associated with cellular telephones.

4.2 Conversational Cues

Due to the limitations of speech recognition technology and privacy concerns, the CNA must only depend on the speech of the user. Most other projects which have tried similar scheduling recognition tasks assume that the complete dialog is available [Stede et al. 1998; Busemann et al. 1997]. Therefore one might believe that it is difficult to obtain all the required scheduling information from the utterances of the user. However, the user can assist the agent by repeating key scheduling times and important points that another speaker has suggested. Since repeating what another person has said is a standard conversational custom for confirming understanding, few people realize that the user is repeating the appointment for the benefit of his scheduling agent. Note that some of the example conversations in Section 3.1 exhibit this behavior. In practice, even experienced wearable computer researchers are sometimes unaware when this "repeating back" behavior is being used to communicate with the wearable user's agent.

"Dual-use" speech is already known to the public in limited forms. For example, while on the telephone with a colleague, a lawyer might say "My assistant Amy will send you the Howard case insurance forms today." This utterance both informs the colleague of the lawyer's intention and provides Amy, who was in the lawyer's office at the time, the specifics needed to fulfill the lawyer's instructions without further interaction. A similar style of telephone conversation can also be seen when a speaker is both communicating with local participants and providing contextual information for the benefit of a remote listener. For example, "Bob needs to take notes on this, so let's give him a second to get his files from his briefcase. OK, he's back at the speaker-phone, so let's think about ..."

4.3 Social Acceptance

Conversational partners may also be distracted by the presence of the user's wearable computer during social discourse. Anecdotally, over the course of ten years wearing machines of various sizes, we find that colleagues do not notice the presence or absence of the equipment for a given interaction within about two weeks of working with a wearable user [Starner 1999]. Even so, less obtrusive technology could lead to faster adoption in the future.

An encouraging development is that displays can now be fitted inside the lenses of a pair of prescription eyeglasses. Figure 2 demonstrates an early prototype of such a device. In the future, one can imagine an entire wearable computer capable of speech recognition embedded in the ear-piece of the eyeglasses or mounted behind the ear like a hearing aid.

Even with unobtrusive wearable technology, one might argue that the use of the machine may interfere with normal social discourse. Society adapts to new technology in conversational practices, as evidenced by the effects of the cellular phone and the PDA. However, minimizing initial social artifacts is one of the main design goals of the next prototype, which will be described in Section 5.

4.4 Cognitive Load

Currently the CNA requires a very specific vocabulary and grammar which we are still refining. As mobile speech recognition technology improves, these requirements will be relaxed. One can imagine an effort similar to the DARPA Airline Travel Information Service (ATIS) task where researchers try to capture the "natural" vocabulary and grammar related to a specific task and then create a system that allows seemingly unlimited interaction while still being specifically tuned to the task [Levin et al. 2000]. Unfortunately, CNA users would still need to formulate their speech to provide enough contextual information to drive the interface (i.e. at least utter the appointment date and time).

Can a CNA user knowingly limit his vocabulary and grammar while in a conversation without an additional cognitive load that inhibits the actual task of scheduling? Does the current push-to-talk interface create an additional cognitive load that is inappropriate? Is the user able to scan the information on the display while maintaining the conversation? These questions will require study but would appear to be a fruitful area of research with applicability to other problems. Already, domains such air traffic control and military communications require variants of push-to-talk for communicating with other participants and require participants to control the vocabulary and structure of their conversation in real-time for clarity. In addition, HUDs are becoming widely used for military applications, where distraction from the primary task is detrimental.

The grammar of the current prototypes is informed by scheduling dialogs from our calendar user study, but with more study on the speech used for the mobile scheduling domain, a socially appropriate and easily memorizable grammar may be chosen for use with limited speech recognition capabilities.

4.5 Limits of Mobile Speech Recognition

Even though the CNA uses push-to-talk and a limited vocabulary and grammar, speech recognition errors will occur. The user can recover from the errors through the Twiddler keyboard with its integrated mouse. Suppose, for example, in mid-August the user said "Let me see if I can meet in the first week of September." and the CNA moved the calendar to the first week of December. Since the usual desktop interface is still available to the user, the user can excuse himself to his conversational partner ("Hold on, let me look up that date.") while manually navigating the interface back to the first week in September. A simple refinement is an "undo" button located in the vicinity of the push-to-talk button that would return the calendar to the previous position before the last utterance was parsed. In this way, the user can avoid excessive navigation cost, and in the worst case the system is roughly equivalent to a desktop or PDA system. A further refinement is a "re-parse" button, which instructs the CNA to perform an undo and then attempt another parsing parsing of the utterance. Such a button can be pressed repeatedly by the user if desired. With experience, the user should be able to judge when to attempt a re-parse and when to navigate the calendar manually.

5. DIALOG TABS

While timing does seem to correlate with the probability that a device will be used as shown in Section 2.2.1.1, it is clearly not the only factor. Section 2.2.2 suggests that the subjects may use a buffer (such as their memory or scrap paper) when mobile to delay the burden of dealing with their primary scheduling device. Postponement may help reduce retrieval time and peak cognitive load, enable entering appointments in batch, or delay commitment until potentially conflicting appointment dates are fixed.

A prototype system called "Dialog Tabs" was designed to explore how audio capture and speech recognition might be used to aid postponing the processing of appointments. The system is designed for low attentional demands during conversation as well as fast access for when the user wants to process the appointment.

5.1 Dialog Tabs Operation

The current Dialog Tabs prototype requires a push-to-talk interface similar to the CNA. Again the utterances allowed while pushing the button are restricted by a predefined grammar, providing various ways to indicate appointment dates.

In contrast with the CNA, the user does not navigate his calendar during the conversation, but records his speech for later processing. When the user speaks a date during normal conversation while holding the push-to-talk button, the system not only tries to recognize the speech, but also stores a predefined number of seconds of audio from both before and after the utterance.

In addition visual feedback is generated by showing a special type of widget, called a Dialog Tab. It is non-modal and appears as a thin vertical bar on an edge of the HUD without shifting focus from the user's current application (Figure 10). As new appointment scheduling events occur, tabs are stacked vertically in order of arrival. The most recent tab is the tallest, covering twice as much screen space as the next most recent tab. The third most recent tab has half the height of the second most recent tab, and so on for as many tabs as displayed [3].

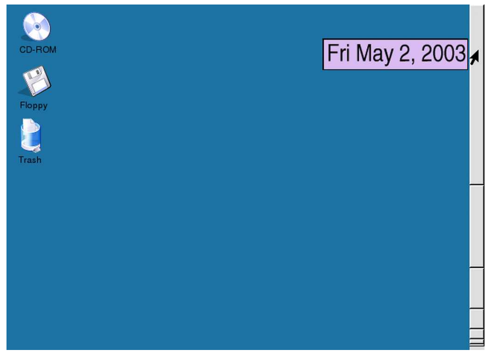

Fig. 10. Dialog Tabs display unobtrusively on the right side of a desktop.

During the day, Dialog Tabs may queue up, but the user does not need to process them until he has the time and inclination to do so. The tabs provide a constant reminder of the appointments that need to be processed; thus the user can postpone processing the events without fear of forgetting them.

The Dialog Tabs require little desktop space and are designed to be minimally distracting even when they appear on a head-up display during conversation. They are also developed so that they can be processed quickly (e.g. as the user walks to his next appointment). By placing Dialog Tabs on the edge of the screen and taking advantage of Fitt's law [Walker and Smelcer 1990], this interface allows the user to process tabs quickly and efficiently even in poor motor control situations such as while walking [Lyons and Starner 2001]. Hovering the mouse over a tab displays the date discussed in the user's conversation as parsed by the speech recognition engine. The hovering strategy is unobtrusive; the user is only presented with detailed information upon request. This concept is similar to Rhodes' "ramping interface" [Rhodes 2000] where the agent progressively discloses more information to the user as he indicates deepening interest with progressively more involved interactions. At any time the user can abandon an interaction quickly and proceed with another task.

If the user wants to access more details about a specific appointment, he left-clicks on the appropriate tab and a modal dialog box appears (Figure 11). The dialog box shows a visual representation of the recorded audio. The sentence that was recognized by the speech recognition engine is displayed below the audio waveform. Hovering the mouse over a word in the sentence highlights the corresponding section in the audio visualization and vice versa, enabling rapid indexing similar to the SCANMail system [Whittaker et al. 2002]. Clicking on a word or part of the waveform results in audio playback starting at the corresponding position. The user can also replay the entire audio sample or modify the date if it was recognized

[3]In theory, a high number of dialog tabs can be accommodated, but since it is difficult to perceive a widget with a height of only a few pixels (especially on a head-up display) we limited the maximum number of tabs to eleven. The twelfth tab is automatically hidden.

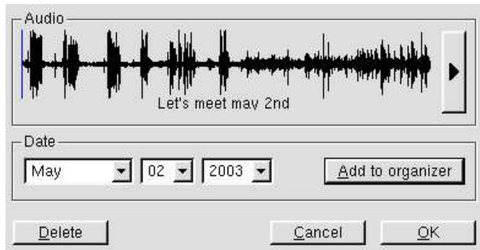

Fig. 11. The Dialog Tabs dialog box allows the user to listen to portions of the audio file by selecting a word or section of the displayed waveform.

incorrectly. The appointment can be discarded or entered into the user's calendar application with a single key press.

The Dialog Tab can also be discarded or added to the calendar application directly via a pop-up menu that appears when right-clicking the on dialog tab, bypassing the modal dialog box and thereby enabling even faster processing.

5.2 Dialog Tabs Performance

To date, no controlled environment timing experiments have been performed on the Dialog Tabs prototype. However, in this section we hypothesize that Dialog Tabs could outperform other postponement scheduling mechanisms like scrap paper.

Similar to the CNA timing test (Section 3.2), memory users are not considered, and the wearable computer is assumed to be instantly accessible by its user, thereby eliminating retrieval time. In our calendar user study, scrap paper takes an average of 17.8 seconds to retrieve. In addition, writing down temporary appointment information on scrap paper during the conversation takes the subjects in our user study on average 18.1 seconds. In contrast Dialog Tabs users only utter a sentence similar to the one they would have spoken anyway, while pushing the push-to-talk button.

For scrap paper users, another scheduling device has to be retrieved and navigated, and the appointment information must be copied from the temporary device during reconciliation. The duration of these actions is dependent on the device the scrap paper users are reconciling with, but clearly the time is not negligible. The total interaction time is probably almost twice the time of entering the appointment in a more permanent device once. With dialog Tabs, in contrast, navigating to a correctly recognized appointment date and entering basic appointment information in a calendar application is done by selecting the appropriate Dialog Tab (usually the top one) and pushing a single button. If the uttered date was not correctly recognized and the user does not recollect the appointment information, a quick correction is possible after replaying the appropriately indexed part of the recorded audio.

5.3 Dialog Tabs Discussion

Inherent to the use of buffering is the inability to identify scheduling conflicts before reconciliation. Therefore an appointment can usually not be confirmed or rejected during a scheduling conversation. This strategy is inefficient when multiple participants are involved in scheduling or when participants have a busy schedule, because of the large number of potential conflicts [Palen 1999]. In these circumstances, underlying changes in people's calendar while buffering could lead to the need to reschedule completely. Dialog Tabs may try to assist by automatically searching the user's calendar and highlighting the Dialog Tab to warn the user on a potential scheduling conflict. Unfortunately, such a system would be more dependent on speech recognition accuracy and therefore occasionally distract the user unnecessarily. Fortunately, the concept of tentative appointments is well accepted in the world of PDAs, which require synchronizing with desktops to receive the most up-to-date version of the schedule. In many cases then, the potential negative results of postponement with Dialog Tabs is no worse than current systems that are currently socially accepted.

Dialog Tabs has two distinct advantages over memory and scrap paper. First, the user always has an instant overview of appointments that still need to be processed, whereas scrap paper can get lost or forgotten, and memory is transient. Second, by recording part of the conversation, additional cues are presented to the user. Journal studies by [Wagenaar 1985] suggest that these cues help the user to recall sufficient information to process partial scheduling conversations at the end of the day. Even when test subjects could not independently recall an event, providing more information about the event seems to allow subjects to recall progressively more independent facts about the event.

Even though Dialog Tabs is as prone to speech recognition errors as the CNA, the impact of incorrectly recognized speech is considerably smaller, since the progress of the scheduling conversation is not dependent on the recognized speech. The user can not get "lost" in his calendar, which eliminates the need for an "undo" button or manual interaction with the agent during the conversation. Even though extra effort is required to modify incorrectly recognized information upon reconciliation, Dialog Tabs requires less user attention during the conversation than the CNA because information on the HUD needs not be processed immediately. In addition, failing to use the predefined grammar has no immediate consequences. These properties enable Dialog Tabs to overcome potential social artifacts that might affect the CNA, such as stilted speech or wandering eye gaze.

6. CONCLUSION AND FUTURE DIRECTIONS

In this paper, some of the promises and challenges for conversational speech-based systems on wearable computers have been discussed. We have chosen the task of appointment scheduling to examine in detail. An extensive user study of current appointment scheduling techniques was performed to inform our research. The results of this study suggest that numerous factors influence the choice of scheduling mechanism, and we used these observations for the design of two prototype systems, the Calendar Navigator Agent (CNA) and Dialog Tabs.

The CNA monitors appointment scheduling conversations and displays calendar information based on the progress of that conversation. In order to constrain the speech recognition problem, the user exploits a variant of push-to-talk and maintains a grammar of socially applicable but narrowly defined utterances. A pilot study showed that the CNA outperformed more conventional scheduling devices on speed in a constrained environment.

Legal, social, psychological, and technical issues on the CNA were discussed, and an alternative scheduling agent, Dialog Tabs, was developed to explore different design attributes. Dialog Tabs captures appointment scheduling conversations for later processing by the user, and therefore speech recognition errors have no direct effect during the conversation. It provides a reminder about a possible appointment without forcing the user to interact with a distracting calendar interface during a conversation at the expense of direct appointment closure and increased total interaction time.

In the future, we wish to expand testing for both the CNA and Dialog Tabs to a mobile environment with multiple users. The inclusion of a key-phrase spotter might ease the push-to-talk constraint for both prototypes. The use of Dialog Tabs as a more general reminder system for wearable computers can also be explored.

Since the sample population of the calendar user study is currently restricted to technical students, repeating the study in an area with a higher concentration of busy professionals who may use scheduling aids more often would provide a broader insight into scheduling behavior.

7. ACKNOWLEDGMENTS

The authors would like to thank James Fusia, Brad Singletary, Helene Brashear, Amy Hurst, Mel Eriksen, and the reviewers for their help in this project. Some of this paper was written on CharmIT Pro wearable computers. This material is supported, in part, by National Science Foundation Career Grant #0093291. This publication is also supported, in part, by the Rehabilitation Engineering Research Center on Mobile Wireless Technologies for Persons with Disabilities, which is funded by the National Institute on Disability and Rehabilitation Research of the U.S. Department of Education under grant number #H133E010804.

REFERENCES

Abowd, G., Atkeson, C., Brotherton, J., Enqvist, T., Gulley, P., and LeMon, J. 1998. Investigating the capture, integration and access problem of ubiquitous computing in an educational setting. In *CHI*. ACM, New York.

Blackwood, W. 1997. *Tactical Display for Soldiers*. National Academy of Sciences, Washington, D.C.

Busemann, S., Declerck, T., Diagne, A. K., Dini, L., Klein, J., and Schmeier, S. 1997. Natural language dialogue service for appointment scheduling agents. Tech. Rep. RR-97-02, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH.

Cohen, P. and Oviatt, S. 1995. The role of voice input for human-machine communication. In *Proceedings of the National Academy of Sciences*. Vol. 92. 9921–9927.

Collins, C., Scadden, L., and Alden, A. 1977. Mobility studies with a tactile imaging device. In *Fourth Conf. on Systems and Devices for the disabled*. Seattle, WA.

Danis, C., Comerford, L., Janke, E., Davies, K., DeVries, J., and Bertrand, A. 1994. Storywriter: A speech oriented editor. In *CHI*. ACM, New York, 277–278.

Hemphill, C. T., Godfrey, J. J., and Doddington, G. R. 1990. The ATIS spoken language systems pilot corpus. In *Proc. of the Speech and Natural Language Workshop*. Hidden Valley, PA, 96–101.

James, F. 2002. Panel: Getting real about speech: Overdue or overhyped. In *CHI*. ACM, New York.

JUNQUA, J. 1993. The lombard reflex and its role on human listeners and automatic speech recognizer. *J. Acoustic. Soc. Amer. 93*, 510–524.

KARAT, C., HALVERSON, C., HORN, D., AND KARAT, J. 1999. Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *CHI*. ACM, New York, 568–572.

KARL, L., PETTEY, M., AND SHNEIDERMAN, B. 1993. Speech versus mouse commands for word processing applications: An empirical evaluation. *Intl. J. Man-Machine Studies 39,* 4, 667–687.

KUBALA, F., ANASTASAKOS, A., MAKHOUL, J., NGUYEN, L., SCHWARTZ, R., AND ZAVALIAGKOS, G. 1994. Comparative experiments on large vocabulary speech recognition. In *ICASSP*. Adelaide, Australia.

LEVIN, E., PIERACCINI, R., AND ECKERT, W. 2000. A stochastic model of human-machine interaction for learning dialog strategies. *Trans. on Speech and Audio Processing 8,* 1, 11–23.

LYONS, K. AND STARNER, T. 2001. Mobile capture for wearable computer usability testing. In *Intl. Symp. on Wearable Computers*. IEEE, Zurich, Switzerland.

NAJJAR, L., THOMPSON, C., AND OCKERMAN, J. 1997. A wearable computer for quality assurance inspectors in a food processing plant. In *IEEE Intl. Symp. on Wearable Computers*. IEEE Computer Society.

NEWMAN, D. 2000. Speech interfaces that require less human memory. *Speech Technology*.

OCKERMAN, J., NAJJAR, L., AND THOMPSON, C. 1997. Wearable computers for performance support. In *IEEE Intl. Symp. on Wearable Computers*. IEEE Computer Society.

OLSEN, D. R. AND PEACHEY, J. R. 2002. Query-by-critique: Spoken language access to large lists. In *UIST*. ACM, New York.

OVIATT, S. 1999. Ten myths of multimodal interaction. *Communications of the ACM 42,* 11, 74–81.

PALEN, L. 1999. Social, individual and technological issues for groupware calendar systems. In *CHI*. 17–24.

RHODES, B. J. 2000. Just-in-time information retrieval. Ph.D. thesis, MIT Media Laboratory, Cambridge, MA.

ROSS, D. AND BLASCH, B. 2000. Wearable interfaces for orientation and wayfinding. In *ACM conference on Assistive Technologies*. 193–200.

ROY, D., SAWHNEY, N., SCHMANDT, C., AND PENTLAND, A. 1997. Wearable audio computing: A survey of interaction techniques. Tech. rep., MIT Media Laboratory.

SCHACTER, D. 2001. *The Seven Sins of Memory*. Houghton Mifflin, Boston.

SCHMANDT, C. 1994a. *Voice Communication with Computers*. Van Nostrand Reinhold, New York.

SCHMANDT, C. 1994b. *Voice Communication with Computers: Conversational Systems*. Van Nostrand Reinhold, New York.

SHNEIDERMAN, B. 2000. The limits of speech recognition. *Communications of the ACM 43,* 9 (September).

SMAILAGIC, A. AND SIEWIOREK, D. 1994. The CMU mobile computers: A new generation of computer systems. In *COMPCON '94*. IEEE Computer Society Press, 467–473.

STARNER, T. 1999. Wearable computing and context awareness. Ph.D. thesis, MIT Media Laboratory, Cambridge, MA.

STARNER, T., WEAVER, J., AND PENTLAND, A. 1998. Real-time American Sign Language recognition using desk and wearable computer-based video. *IEEE Trans. Patt. Analy. and Mach. Intell. 20,* 12 (December).

STEDE, M., HAAS, S., AND KÜSSNER, U. 1998. Tracking and understanding temporal descriptions in dialogue. Verbmobil-Report 232, Technische Universität Berlin. October.

STEIN, R., FERRERO, S., HETFIELD, M., QUINN, A., AND KRICHEVER, M. 1998. Development of a commercially successful wearable data collection system. In *IEEE Intl. Symp. on Wearable Computers*. IEEE Computer Society.

STIFELMAN, L. 1996. Augmenting real-world objects. In *CHI*. ACM, New York.

STIFELMAN, L., ARONS, B., SCHMANDT, C., AND HULTEEN, E. 1993. Voicenotes: A speech interface for a hand-held voice notetaker. In *CHI*. ACM, New York, 179–186.

Strubb, H., Johnson, K., Allen, A., Bellotti, V., and Starner, T. 1998. Privacy, wearable computers, and recording technology. Panel discussion, The Second International Symposium on Wearable Computers, October 19–20, 1998, Pittsburgh, PA.

Upton, M. 1968. Wearable eyeglass speechreading aid. *American Annals of the Deaf 113*, 222–229.

Wagenaar, W. 1985. My memory: A study of autobiographical memory over six years. *Cognitive Psychology* 18, 225–252.

Walker, M., Wright, J., and Langkilde, I. 2000. Using natural language processing and discourse features to identify understanding errors in a spoken dialogue system. In *Proc. 17th International Conf. on Machine Learning*. Morgan Kaufmann, San Francisco, CA, 1111–1118.

Walker, N. and Smelcer, J. 1990. A comparison of selection time from walking and bar menus. In *Proceedings of CHI'90*. ACM, Addison-Wesley, New York, 221–225.

Whittaker, S., Hirschberg, J., Amento, B., Stark, L., Bacchiani, M., Isenhour, P., Stead, L., Zamchick, G., and Rosenberg, A. 2002. Scanmail: a voicemail interface that makes speech browsable, readable and searchable. In *CHI*. ACM Press, New York, 275–282.

Whittaker, S., Hyland, P., and Wiley, M. 1994. Filochat: Handwritten notes provide access to recorded conversations. In *CHI*. ACM Press, New York, 271–276.

Wickens, C. 1984. *Varieties of Attention*. Academic Press, New York, Chapter Processing resources in attention.

Wilcox, L., Schilit, B., and Sawhney, N. 1997. Dynomite: A dynamically organized ink and audio notebook. In *CHI*. ACM, New York, 186–193.

Yankelovich, N., Levow, G., and Marx, M. 1995. Designing SpeechActs: Issues in speech user interfaces. In *CHI*. ACM, New York, 568–572.

Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Haze, T., and Hetherington, L. 2000. Jupiter: A telephone-based conversational interface for weather information. *Trans. on Speech and Audio Processing 8,* 1, 85–96.