Providing Support for Mobile Calendaring Conversations: A Wizard of Oz Evaluation of Dual–Purpose Speech

Kent Lyons, Christopher Skeels, Thad Starner
College of Computing and GVU Center
Georgia Institute of Technology
Atlanta, GA 30332-0280 USA
{kent, cskeels, thad}@cc.gatech.edu

ABSTRACT

We present a Wizard of Oz evaluation of dual–purpose speech, a technique designed to provide support during a face–to–face conversation by leveraging a user's conversational speech for input. With a dual–purpose speech interaction, the user's speech is meaningful in the context of a human–to–human conversation while providing useful input to a computer. For our experiment, we evaluate the ability to schedule appointments with our calendaring application, the Calendar Navigator Agent. We examine the relative difference between using speech for input compared to traditional pen input on a PDA. We found that speech is more direct and our participants can use their conversational speech for computer input. In doing so, we reduce the manual input needed to operate a PDA while engaged in a calendaring conversation.

Categories and Subject Descriptors: H.5.2 [User Interfaces]: Voice I/O, Natural Language, Input devices and strategies.

General Terms: Human Factors, Experimentation.

Keywords: Speech user interfaces, dual–purpose speech, mobile computing.

1. INTRODUCTION

Much of our lives is spent communicating with others. A study of office workers found that 25–85% of their time at work was spent in interpersonal communication [4]. Increasingly, our interactions are in mobile settings; for two office workers, Whittaker et al. found that 17% of their total work day was spent in conversations while "roaming" or away from the desk [6].

In our previous work [3], we developed dual–purpose speech, an interaction technique designed to leverage a user's conversational speech and reduce manual input. A dual–purpose speech interaction is one where the speech serves two roles. First, it is socially appropriate and meaningful in the context of a human–to–human conversation. Second, the speech provides useful input to a computer. A dual–purpose speech application maintains the privacy of others by only listening to the user's speech and provides beneficial services during the conversation. Since so many of our conversations happen while away from the desk, our dual–purpose speech applications are designed to run on mobile devices such as PDAs or wearable computers. By using a mobile device, we enable the use of our applications during the serendipitous face–to–face conversations that occur throughout the day. In this paper, we present an experiment which evaluates dual–purpose speech.

MobileHCI'05, September 19–22, 2005, Salzburg, Austria.

1.1 The Calendar Navigator Agent

We have found previously that traditional PDA interfaces can consume a significant amount of time during a calendaring conversation[5]. The Calendar Navigator Agent (CNA) is an application which automatically navigates a user's calendar based on socially appropriate speech used while scheduling appointments. The CNA is a traditional calendar application that has been augmented to utilize speech during a social interaction. The goal is to allow user interaction with the calendar that minimally disrupts the scheduling conversation.

The Calendar Navigator Agent has a graphical interface similar to other scheduling applications found on a PDA (Figure 1). As the user proceeds with a conversation, he can press a push–to–talk button to activate the speech recognition engine (indicated with bold below). The speech fragment is then processed by the speech recognition system and specific keywords such as "next week" or "Monday" are recognized. The CNA uses the speech recognition results to perform the relevant calendaring actions. If an error is made and an improper action is performed, the user can press a single button to undo the last command. For example, the following hallway conversation between the CNA user (P) and another person (R) is representative of speech the CNA supports:

R: "Can we meet next week?"
P: "Sure, **how about next Tuesday?**"
The CNA shows next Tuesday.
R: "That would work. How about 2pm?"
P: "Ok, **I'll see you at 2**"
The CNA creates the appointment at 2pm.

Our experiment is designed to determine if novice participants can remain engaged in a dialog while at the same time use their speech to control a computer. Our study is constructed such that novice dual–purpose speech users schedule a sequence of appointments with a researcher using our calendaring application on a PDA we provide. For

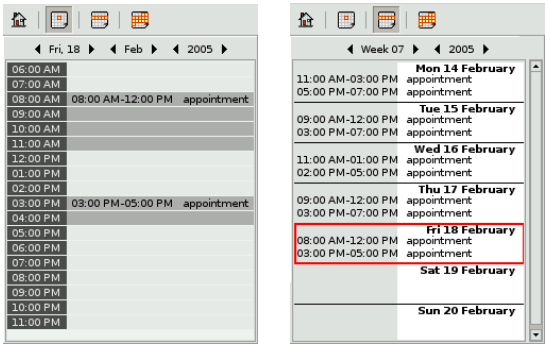

Figure 1: The day and week views of the calendar.

each appointment, the researcher initiates a calendaring discussion with a goal, e.g. "Can we meet next week?" The participant then navigates the calendar on the PDA while negotiating a suitable time with the researcher. After a time is agreed upon, an appointment is created.

2. EXPERIMENT DESIGN

For our experiment, we compare a dual–purpose speech condition and a control condition using pen input. Both conditions are performed by every participant resulting in a within–subjects design, and the order of conditions is counterbalanced across participants. For the pen input method, the participant uses the PDA stylus to navigate the calendaring application. In the speech condition, the participant uses dual–purpose speech. During each condition, the participant performs twenty appointment creation trials with the given input method. Each trial is untimed, and the entire experiment takes approximately 45 minutes per participant to complete.

2.1 Wizard of Oz

We are utilizing the Wizard of Oz technique in place of an automatic speech recognition system. A second researcher, known as the wizard, simulates the recognition and semantic processing of the participants' speech [1]. This technique enables flexibility in the language we accept with our dual–purpose speech application and enables our participants to use their preexisting scheduling language.

2.2 Trials

The researcher scheduled twenty unique appointments with the participants for each condition. He initiated all of the scheduling dialogs with a phrase similar to "Can we meet..." That phrase was then followed by one of 40 different predefined scheduling goals such as "on a Wednesday morning?" "February 17th?" or "the week of March 21st?" Each trial was designed to simulate a different scheduling dialog that may occur opportunistically during daily conversations.

2.3 Participants

We recruited twenty participants from our Institute and all were compensated $10 for their time. Our participants ranged from 19 to 39 years old and had a mean age of 25.9 years ($SD = 5.8$). Fifteen participants were male. Nineteen of our twenty participants reported that they owned a cell phone, while six indicated they owned a PDA. Our participants reported that they spent an average of 14.4 hours per week in scheduled meetings or classes ($SD = 7.0$). Fourteen of the participants indicated they had prior experience with speech recognition, many with automated phone response systems.

2.4 Procedure

The experiment began with the researcher presenting an overview of the experiment. Participants filled out consent and demographic forms and were given written instructions describing the calendar application.

For the pen condition, the participant navigated the calendar using the PDA stylus for input. The session began with some practice interactions designed to familiarize the participant with the pen input mechanism. The researcher instructed the participant to navigate through a sequence of days and weeks, and after completing the predefined navigations and appointment creations, the participant was instructed to try a few interactions of his or her choosing. Once the participant was satisfied, the trials began.

For the dual–purpose speech condition, the participant navigated the calendar using speech input. Again, the session began with practice. The researcher instructed the participant how to use the push–to–talk and undo buttons and asked him to perform a simple navigation using speech. Next, the researcher described dual–purpose speech and instructed the participant that he could use a single utterance to fill two roles. The researcher then stepped the participant through a simple dialog that showed how to use dual–purpose speech and how the speech affected the calendar. After the predefined navigations were complete, the participant used speech input to control the calendar at his own discretion. Once the participant was comfortable, practice ended, and trials began.

2.5 Software and Equipment

While we intend the CNA to be used on a mobile computer during the course of everyday activities, we constrained the study to a stationary test and conducted the experiment in a usability laboratory. The participant and researcher sat at a table facing each other. The participant was provided with our calendaring software running on an iPaq PDA which was held in his or her hand in a comfortable position. The wizard was located out of sight in an adjacent room, and the participants were not informed that the speech was being recognized by a wizard.

The software used was a modified version of the calendar from the GPE Palmtop Environment[1] (Figure 1), a collection of open source software which runs on Linux and uses the X Window System. The calendar was modified so one of the buttons on the front of the PDA acted as an undo for both conditions, and another button was used for the push–to–talk functionality used during speech input. The software records an event log of the user's interaction with the PDA and audio from two microphones. The first microphone is a headset which only recorded the participants' speech, and the second microphone was placed on the desk and recorded the entire conversation.

As discussed above, we did not use a real speech recognition system and instead simulated one with a wizard. We

[1]http://gpe.handhelds.org/

implemented this functionality by routing the captured audio of the participant's voice to the wizard in the adjacent room. In normal operation, the wizard's audio is muted, and he cannot hear any of the appointment dialog. In the speech condition, the push–to–talk button un–mutes the wizard's audio. As a result, the wizard hears only the portions of the participant's speech when the button is depressed, which in turn enables him to simulate a speech recognition system. The interface used by the wizard is also an extension of the GPE calendar which was modified with additional windows to allow quick navigation and appointment creation based on the user's speech and ran on a desktop computer.

The software ran on the researcher's computer and was shared with the iPaq and wizard's computer with VNC[2]. VNC is an application that can export the graphical user interface of a program to remote computers and allow remote users to interact with that program. For this experiment, we used VNC to export the main calendar window to the PDA so that the participant saw what appeared to be a traditional PDA calendar. The rest of the application was used by the wizard and researcher to run the experiment.

3. FINDINGS

For each of our twenty participants, we collected twenty appointment dialogs using speech for input and twenty more using pen. In total, we have 800 different calendaring conversations.

3.1 Comparing Pen and Speech Input

We are interested in the relative performance differences between our two input conditions. In general, we found small differences in our metrics which are summarized in Table 1. Also presented are the p–values from unpaired t–tests used to compare the two populations. The conversations were slightly shorter for the pen input condition, taking on average approximately 3 seconds less time ($M_p = 20.9s, M_s = 17.8s$). In contrast, excluding the researcher's side of the conversation and examining only the duration of each individual turn during the dialog yields no significant difference ($p = 0.385$). Comparing the cumulative duration for all of the participants' turns during a task shows the conversation is 2.2s shorter when using pen ($M_p = 13.2$) relative to speech ($M_s = 15.4$). Likewise, comparing the number of turns per trial shows that there were slightly more turns taken using speech ($M_s = 2.9$) than with pen ($M_p = 2.5$). In summary, the participants held the floor slightly longer and more frequently when using speech for input, and the overall conversation took a few seconds longer. Together the data imply that the participants were talking more in the speech condition.

One of the strongest differences between the conditions is the number of navigations used for an appointment. We use the term "navigation" to denote any change in the state of the application. For instance, switching from day view to week view, advancing a week, or selecting a particular day are considered as navigations. With speech, our participants performed $M_s = 1.3$ navigations per appointment dialog. In contrast, the pen users performed on average an extra two navigations during each conversation ($M_p = 3.3$). While speech is not necessarily faster, it has the advantage of being more direct. The speech user can navigate to her intended location in the calendar in fewer steps.

[2]http://www.realvnc.com/

Metric	Speech	Pen	P-value
Duration of conversation	20.9s	17.8s	< 0.001
Duration of turn	5.4s	5.2s	0.385
Cumulative dur. of turns	15.4s	13.2s	< 0.001
Number of turns	2.9	2.5	< 0.001
Number of navigations	1.3	3.3	< 0.001

Table 1: Results from pen and speech conditions.

We were also interested in the subjective differences between our two methods. We used the NASA Task Load Index (TLX) questionnaire to obtain a measure of the workload imposed by our tasks [2]. There was no effect for overall workload, and likewise most of the subcomponents showed no effect between our conditions. The two dimensions that do have a significant difference are performance and physical demand. Participants rated their performance to be better using pen compared to speech ($M_p = 5.6$ and $M_s = 13.1$ respectively out of 100), and not surprisingly, participants rated the speech input method to be less physically demanding than the pen input method ($M_s = 2.4$ and $M_p = 10.1$ respectively out of 100). We administered a Likert scale questionnaire on the naturalness and flow of the calendaring conversations. This questionnaire revealed no statistically significant differences between the speech and pen conditions for any of our five questions. This result implies that our participants did not think that one input method or the other caused more disruption to the flow of the conversation.

Finally, we were interested in the use of the push–to–talk button and characterizing the delay of speech processing from the wizard. On average, our participants held the push–to–talk button for 1.3s ($SD = 0.67$). Our wizard took an average of 1.5s ($SD = 0.86$) to complete an action once the participant released the push–to–talk button. Changing the delay in speech processing could help increase the performance of using speech for input. As the speed of mobile computers increases the time needed to process the user's speech similarly decreases. With a fast enough computer, eventually this delay could be negligible.

3.2 Use of Dual–Purpose Speech

In the previous section we detailed the relative performance of pen and speech input. Now we focus on some of the more qualitative aspects of dual–purpose speech uncovered by our study. First is the nature of the speech used. The idea behind dual–purpose speech is that the user can create an utterance that very naturally fits into the flow of the conversation with her partner but at the same time provides input to the computer. Our participants had varying degrees of success in using speech as we intended. At one extreme, one participant used very structured speech that was directed primarily at the computer. For instance, to navigate the calendar, he would use a phrase such as "Show me the 23rd" or "Create an appointment at 2pm on Tuesday the 5th." After the experiment, the participant indicated that he was not sure what the speech recognition system could understand. Therefore, he intentionally decided not to vary his speech during the conversation in fear of the system not understanding him. It is possible that this participant's

prior knowledge of the limitations of speech recognition led to this behavior.

In contrast, the rest of our participants were much more fluid with their use of speech during most of the conversations. For instance, they might say "Let me check the 23rd" or "Would Wednesday work?" While the participants were often explicitly addressing the computer, the speech also fit into the context of the conversation. The primary exception (besides the one participant described above) is at the beginning of the conversation. As described in our experimental design, the researcher initiated all of the scheduling dialogs. As a result, the participant would often echo the exact same phrase so that the computer could act upon it. For instance, the researcher would say: "Can we meet on February 17th" and the participant would echo back "February 17th." Occasionally participants would change their intonation and echo the phrase as a question seeking confirmation from the researcher that they heard correctly, but given the large number of trials each participant performed, they often did not persist with the strategy of turning the phrase into a question.

In a real calendaring scenario this echoing behavior would likely not be an issue. First, many of the appointments would be initiated by the CNA user and the initial utterance could be used for input. Even for the cases where the user does not initiate, the strategy of echoing for confirmation could be quite useful. Though tedious for our experiment where we scheduled 20 appointments in succession, it might be practical if the appointments were spread out over a larger period of time such as over the course of a day or week. It is also interesting to note that several participants used this confirmation echo during the pen condition, even before using speech for input. Often they would speak more quietly or to themselves while they were using the pen to navigate, suggesting that the strategy is already present for some participants.

As we intended, some of our participants became very adept at constucting phrases that were both appropriate for the scheduling conversation and useable as computer input. In particular, several participants independently developed a strategy that we have named "speculative scheduling" which involves a creative use of the undo button. Originally, we only intended the undo to be used as a way to compensate with errors in the speech recognition which still occasionally happened with the wizard.

At the end of a dialog when a time is decided, the conversation might proceed as follows. The participant would suggest "How about 2 o'clock" and the researcher would respond, "2 works for me." In the speech condition some participants would then just echo "2" again while pressing the push–to–talk button to enter the appointment. Other participants, however, used speculative scheduling. The participant would preemptively enter an appointment by pressing the push–to–talk button while suggesting a time. If the researcher agreed, the appointment was already completed and the dialog was done. If the time was not good, the participant would press the undo button erasing the appointment and then either create a new one using the researcher's suggestion or repeat the speculative scheduling process:

P: "**Would 2 o'clock work?**"
CNA creates an appointment at 2pm.
R: "No, I can't meet then."
P: *Pushes the undo button.*
CNA erases the 2pm appointment.
P: "OK. **How about 4 o'clock?**"
CNA creates an appointment at 4 pm.
R: "That works for me."

By using the undo in this way, participants could create an appointment using truly dual–purpose speech. If they were successful the task would be finished, and if they were not, the cost of removing the appointment was extremely low (a single button press). This strategy is particularly interesting not only because it represents a good example of dual–purpose speech, but also because several participants discovered it independently without any instruction.

4. CONCLUSIONS

Our data suggests that novices can use the Calendar Navigator Agent to control their mobile computer as part of a calendaring conversation in normal discourse. While our speech condition did not show a performance benefit, it did result in a conversation where the participant held the conversational floor longer. Speech input is also much more direct than pen; our participants needed fewer navigations during the scheduling dialog when using speech.

One of the most interesting and unexpected uses of dual–purpose speech was "speculative scheduling" where the participant used dual–purpose speech to create an appointment and used the undo if it did not fit the needs of his partner. While not all of our participants discovered this strategy, overall our novices successfully adopted dual–purpose speech. Even with very little training, our participants quickly determined how to construct their speech so the researcher understood it and the computer could act upon it.

Together, the data from this experiment show that dual–purpose speech is an effective input mechanism and novice users quickly adapt to the technique. By using dual–purpose speech we can reduce the amount of manual input required. Our data indicate that dual–purpose speech applications are a viable mechanism for supporting mobile interaction during the conversations that occur in everyday life.

5. REFERENCES

[1] J. Gould, J. Conti, and T. Hovanyecz. Composing letters with a simulated listening typewriter. *Communications of the ACM*, 26(4), April 1983.

[2] S. G. Hart and L. E. Staveland. *Human mental workload*, chapter Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. North-Holland, 1988.

[3] K. Lyons, C. Skeels, T. Starner, C. M. Snoeck, B. A. Wong, and D. Ashbrook. Augmenting conversations using dual–purpose speech. In *Proceedings of UIST 2004*, 2004.

[4] R. Panko. Managerial communication patterns. *Journal of Organisational Computing*, 1992.

[5] T. Starner, C. M. Snoeck, B. A. Wong, and R. M. McGuire. Use of mobile appointment scheduling devices. In *Proceedings of CHI*. ACM Press, 2004.

[6] S. Whittaker, D. Frohlich, and O. Daly-Jones. Informal workplace communication: what is it like and how might we support it? In *Proceedings of the CHI*, 1994.