Enhancing Support of Conversations by Improving Mobile Computing Input

Kent Lyons
College of Computing and GVU Center
Georgia Institute of Technology
Atlanta, GA 30332-0280 USA
kent@cc.gatech.edu

ABSTRACT

Mobile devices are becoming more commonplace and part of our daily lives. In this work we explore enhancing conversations by improving input for mobile computing. We are researching two ways to enhance mobile input: reusing conversational information through dual–purpose speech and increasing a user's data entry capability with the Twiddler chording keyboard.

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

Additional Keywords and Phrases: Input, mobile computing, speech user interfaces, text entry

INTRODUCTION

Mobile computing is becoming one of the most widely adopted computing technologies. There are currently 1.3 billion mobile phone subscribers and there could be as many as 2 billion by 2007 [1]. The current generation of phones offers substantial computing capabilities and represents just one example of mobile computing becoming increasingly incorporated into everyday life. One very common and important everyday activity is conversation: a study of office workers found that 60–85% of their time at work was spent in interpersonal communication [8]. Conversations commonly occur in mobile settings; Whittaker *et al*. found that 17% of their participants' total work day was spent in conversations while "roaming" or away from the desk [11].

With the huge popularity in mobile computing it is critical that we examine the human–computer interaction issues for these devices and explicitly explore supporting everyday activities. The goal of this work is to enable support of conversations by enhancing mobile computer input. We want to allow a user to remain engaged in her primary role of communicating with her conversational partner while also giving her the ability to enter and manipulate data on a mobile device.

In this work, we explore two methods that improve a user's ability to enter information into a mobile computer in conversational situations. First we examine the Twiddler, a keyboard that has been adopted by the wearable computing community. The Twiddler is a mobile one–handed chording keyboard with a keypad similar to a mobile phone. Our second input method uses dual–purpose speech, a technique designed to leverage a user's conversational speech. A dual–purpose speech interaction is one where speech serves two roles; it is socially appropriate and meaningful in the context of a human–to–human conversation and provides useful input to a computer. A dual–purpose speech application listens to one side of a conversation and provides beneficial services to the user.

Our hypothesis is that we can enhance support of conversation by improving mobile input via two complementary methods:

- By increasing a user's data entry capability with the Twiddler chording keyboard, and
- Through reusing conversational information with dual–purpose speech

We explore this thesis through our contributions which include:

- Research determining the learning rate of the Twiddler and a comparison to the common mobile phone entry method of multi–tap.
- An examination of expert characteristics of Twiddler chording including research on the effects of multi–character chords (MCCs) and limited visual feedback.
- An evaluation of improving novice use of the Twiddler through use of a chording tutorial.
- The input technique of dual–purpose speech and three example applications.
- An evaluation of a dual–purpose speech application designed to uncover implementation requirements of dual–purpose speech applications and the associated design space.

First, we explore increasing the user's data entry capability using the Twiddler keyboard. We present an evaluation in which we determine the learning rate of the Twiddler in comparison to multi–tap. We explore some characteristics of expert typing and present our ongoing work on improving the entry rate for novice users. Next, we introduce a new input method, dual-purpose speech, which reuses conversational information. We present three example applications and a proposed experiment on evaluating the effectiveness of dual–purpose speech in the context of a particular applica-

UIST '04, October 24–27, 2004, Santa Fe, New Mexico, USA
ACM 1-58113-962-4/04/0010

Figure 1: On the left, typing using multi–tap on the Twiddler keypad. On the right, chording with the Twiddler one–handed keyboard typing the letter ‘s’.

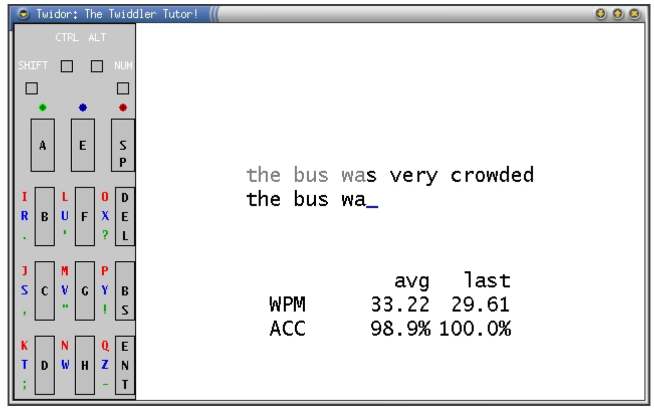

Figure 2: Experimental software showing the chording layout, phrase and statistics.

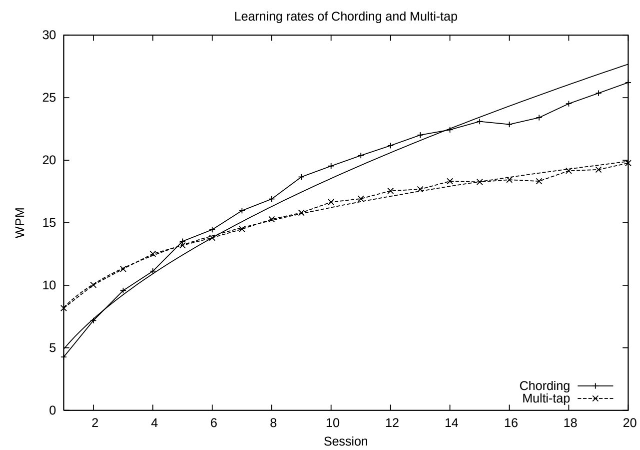

Figure 3: Learning rates and exponential regression curves for multi–tap and chording.

tion, the Calendar Navigator Agent.

MOBILE TEXT ENTRY: THE TWIDDLER

An experienced user of the Twiddler, a one–handed chording keyboard, averages speeds in excess of 60 words per minute (wpm) typing standard test phrases. This fast typing rate coupled with the Twiddler’s 3x4 button design, similar to that of a standard mobile telephone, makes it a potential alternative for text entry on mobile phones. Despite this similarity, there is very little data on the Twiddler’s performance and learnability.

One common mobile phone text entry method is multi–tap, a system in which multiple letters are mapped to the same key and the user presses that key to cycle through the letters until the desired one appears on the screen. Usability studies have found rates of 15.5 wpm [7] and 5.3-10.5 wpm [2] which are far below many other keyboard input devices.

Many wearable computer users [3, 10] type with the HandyKey Twiddler (Figure 1, right). This is a mobile one–handed chording keyboard with a keypad similar to a mobile phone. It has twelve keys arranged in a grid of three columns and four rows on the front of the device. Each row of keys is operated by one of the user’s four fingers. Users hold the device in the palm of their hand like a cup with the keys facing away from their bodies. All five fingers on a hand can be used to type and multiple keys are pressed simultaneously generating a chord.

Text Entry Speeds and Learning Curve

We conducted a longitudinal experiment to compare chording to multi–tap [6]. The subjects participated in 20 sessions over the course of three weeks where each session lasted approximately 45 minutes. Each session consisted of typing text phrases in both conditions and included a 5 minute break. Depending on the condition under test, the testing software presented the participants with the key layout for either multi–tap or chording and statistics of performance. A phrase was presented on the screen and users transcribed the text using the current input method (Figure 2).

For each of our ten participants, we collected approximately 2100 transcribed phrases for a total of 600,000 transcribed characters. The mean entry rates for session one were 8.2 wpm for multi–tap and 4.3 wpm for chording. As sessions continued, the means improved and reached 19.8 wpm for multi–tap and 26.2 wpm for chording by session 20. While both showed improvement, the performance scores for the chording condition rapidly surpassed those of multi–tap (Figure 3). The crossover point in the curves indicates where one condition’s typing rate overtakes the other. In our study, the chording method began with slower speeds but quickly overcame multi–tap. The crossover occurred after the fifth session or after 100 minutes of practice.

Towards Expert Rates

Our next study was designed to confirm the predictions of expert rates made by our original study and explore some characteristics of expert typing. The study continued with a very similar procedure as in our previous experiment [4]. For this study five of our original ten participants agreed to continue, and we resumed testing after a two week intermission. Our procedure was modified to focus on chording. To do

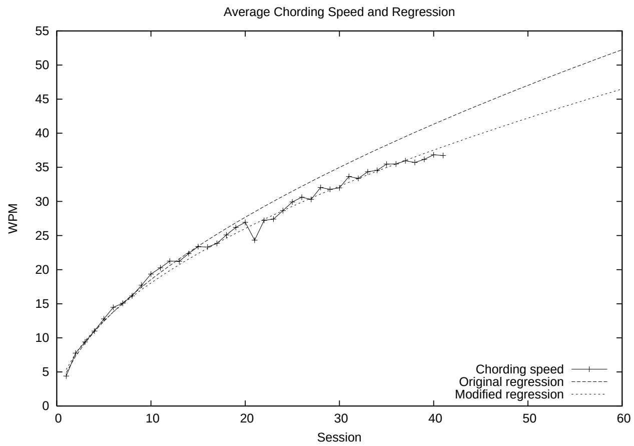

Figure 4: Mean learning rates and regression curves across participants.

this, we replaced the multi–tap condition from our original experiment with a second chording session.

We collected data for approximately 20 additional sessions for a total of 40 sessions or about 13 hours of practice per participant. Figure 4 shows the average typing speed across participants. After 40 sessions the average typing rate for our participants increased to 37.3 wpm. Also plotted is the original regression from our first study and a modified regression based on the new data from our five participants. The curves show that our original regression was slightly optimistic but still closely corresponds to our additional data.

By the end of the study, our five participants achieved an average rate of 47 wpm after approximately 25 hours of practice. One subject achieved a rate of 67 wpm, equivalent to the typing rate of an expert who has been a Twiddler user for ten years.

Blind Typing

In a mobile environment, a user's visual attention may be diverted away from her display while entering text. We conducted an experiment with limited visual feedback with 3 conditions (normal feedback, dots feedback, and blind) using the five participants from our previous study which had been trained up to expert typing rates. Normal feedback displays the text as in our previous experiments. For the dots condition, we display periods for each character typed instead of the transcribed text. This condition is designed to simulate monitoring text typed without being able to actually read the letters, such as looking at a head–up display using only one's peripheral vision. Finally, the blind condition does not show any on-screen indication of what is typed. Based on other work examining typing with mobile phone keypads [9], we hypothesized that reducing the visual feedback would limit our participants' typing rate and accuracy. Surprisingly, changing the visual feedback did not hinder the participants in their typing as expected. In some cases typing and error rates improved with the reduced feedback. This indicates that our participants had successfully learned how to touch type on the Twiddler and were not relying on the display for feedback.

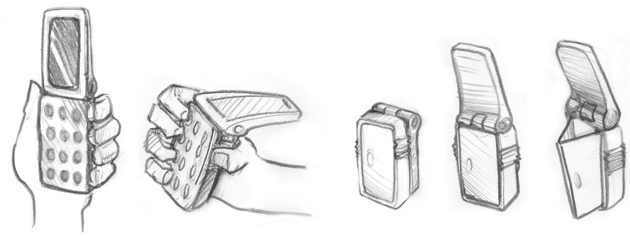

Figure 5: A mobile phone design which incorporates chording capabilities.

Chording as Alternative for Text Entry on Mobile Phones

Both multi–tap and the Twiddler chording method utilize a 12 button 3X4 keypad in a size appropriate for mobile phones. As we have demonstrated through the data collected in these studies, even novice chording users quickly outperform multi–tap typing speeds. One can imagine a mobile phone based on the current Twiddler keyboard; Figure 5 shows one potential design[1]. For messaging or learning to type, the high resolution screen could be used to display a graphical interface similar to that in Figure 2, which would encourage good touch typing. Given the rapid ability to type, this device might enable advanced mobile phone features such as mobile email.

ONGOING WORK: Novice Twiddler Use

The Twiddler offers a faster typing rate than multi–tap after 100 minutes of practice, and with approximately 25 hours the average rate of typing is 47 wpm. Our final experiment will explore how to improve the typing rates of Twiddler novices. Our goal is to have participants typing 8 wpm after 20 minutes, the same rate as multi–tap from our first study.

The first part of this study is designed to determine which combination of two techniques provides the most benefit to novice Twiddler users. We have a 3x2 between–subject design of highlighting (3) and phrase set (2) resulting in 6 conditions. Our software has the capability to highlight the next set of buttons the user is to press. The three highlighting conditions are on, off, and delayed. If a participant cannot remember which buttons to press for a chord, the participant must scan the visual representation to determine which buttons to press. Highlighting simplifies this task by showing which buttons must be pressed to generate the next letter. The delayed condition waits 1.5s for the participant to remember which keys to press then highlights the buttons. The no highlighting offers no additional aid. Each participant will also be assigned to one of two phrase set conditions. The normal phrase set condition utilizes our full phrase set all of the time. Our limited phrase set condition is designed to introduce different categories of chords gradually much like learning the home row first on a desktop QWERTY keyboard.

DUAL–PURPOSE SPEECH

Next, we explore the concept of dual–purpose speech: speech that is socially appropriate in the context of a human–to–human conversation that also provides meaningful input to a computer. We have developed this technique as a way to gain computer support during conversations [5].

[1] Sketches by Stephen Griffin

In a human–to–human conversational situation, it is important that any speech interaction with a computer fit the flow of the conversation. For instance, there are numerous situations where it would be socially inappropriate to talk directly to a computer. By using dual–purpose speech, a person can maintain socially appropriate speech: speech where the language and grammar used fits the conversation. Our dual–purpose speech applications utilize the content from the user's side of the conversation and attempt to minimize disruptions in the flow of conversation by reducing manual interaction with the computer.

We have developed three applications that utilize dual–purpose speech to assist a user in conversational tasks: The Calendar Navigator Agent, DialogTabs, and Speech Courier. The Calendar Navigator Agent (CNA) automatically navigates a user's calendar based on socially appropriate speech used while scheduling appointments. DialogTabs allows a user to postpone cognitive processing of conversational material by providing short–term capture of transient information. Finally, Speech Courier allows the asynchronous delivery of relevant conversational information to a third party.

ONGOING WORK: Dual–Purpose Speech Evaluation

Our final experiment is designed to evaluate the effects of dual–purpose speech in the context of our calendaring application, the Calendar Navigator Agent. Our goal is to determine if dual–purpose speech is an effective way to reuse information from a conversational situation. We are also interested in comparing the use of dual–purpose speech to more traditional interaction methods. We plan on examining the use of the CNA on a personal digital assistant (PDA) and comparing the use of speech to control a calendar relative to using a pen based input method.

Our experiment will be a Wizard of Oz design; the speech recognition and semantic processing will be done by a wizard. This will help us control for speech recognition errors and explore part of the dual–purpose speech design space that is not yet feasible with our current speech recognition implementation.

For our experiment we will have two dual–purpose speech conditions and one control condition. The control condition will be performed by every participant and will involve scheduling a task on a PDA using pen based input. Each participant will also use one of the two speech conditions: restricted language with push–to–talk or unrestricted without push–to–talk. The restricted language with push–to–talk condition represents the capabilities of our current CNA prototype. The downside of this implementation is that it requires the user to perform extra tasks during the conversation to provide input for the computer. The unrestricted language without push–to–talk case is designed to minimize the requirements on the user. Our hypothesis is that the unrestricted case will have a lower cognitive load and the language will be rated as more socially appropriate compared to the restricted case. However, in the unrestricted language case, it is possible that there is not enough content from only the user's side of the conversation for the wizard to successfully determine the appropriate calendaring action.

ACKNOWLEDGMENTS

Thanks to my advisor, Dr. Thad Starner. This material is supported in part by the NIDRR Wireless RERC and NSF Career Grant #0093291.

REFERENCES

1. S. Baker, H. GreenWith, B. Einhorn, M. Ihlwan, A. Reinhardt, J. Greene, and C. Edwards. Big bang! BusinessWeek, June 2004.
2. C. L. James and K. M. Reischel. Text input for mobile devices: comparing model prediction to actual performance. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 365–371. ACM Press, 2001.
3. K. Lyons. Everyday wearable computer use: A case study of an expert user. In *Proceedings of Mobile HCI*, pages 61–75, 2003.
4. K. Lyons, D. Plaisted, and T. Starner. Expert chording text entry on the twiddler one–handed keyboard. In *Proceedings of IEEE International Symposium on Wearable Computing*, 2004.
5. 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.
6. K. Lyons, T. Starner, D. Plaisted, J. Fusia, A. Lyons, A. Drew, and E. Looney. Twiddler typing: One-handed chording text entry for mobile phones. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 2004.
7. I. S. MacKenzie, H. Kober, D. Smith, T. Jones, and E. Skepner. Letterwise: prefix-based disambiguation for mobile text input. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 111–120. ACM Press, 2001.
8. R. Panko. Managerial communication patterns. *Journal of Organisational Computing*, 1992.
9. M. Silfverberg. Using mobile keypads with limited visual feedback: Implications to handheld and wearable devices. In *Proceedings of Mobile HCI 2003*, pages 76–90, 2003.
10. T. Starner. *Wearable Computing and Context Awareness*. PhD thesis, MIT Media Laboratory, Cambridge, MA, May 1999.
11. 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 SIGCHI conference on Human factors in computing systems*, pages 131–137. ACM Press, 1994.