Narrative of the Unigesture User Study

Vibha Sazawal

4 June, 2001

Abstract

This report narrates my experiences while conducting a user study of the Unigesture text entry method. A brief description of the Unigesture method is outlined, followed by the design of the user study. I also describe a Human Subjects application that was filed. The paper briefly discusses results of the user study and concludes with commentary and lessons learned from the entire process.

1 Overview of Unigesture

This section highlights aspects of Unigesture needed to understand the user study. For full details, I refer the reader to [1].

1.1 Introduction

Unigesture is a method for one-handed text entry. It is intended for small devices and handheld computers. To write, the user tilts the device with the same hand that is holding it. Tilt-based UI's are a well-known alternative for devices too small for a keyboard or stylus – handheld devices such as the Hikari [2] and Itsy [3] use tilting to scroll through photos and lists of information. Unigesture applies tilting to a very general-purpose task – selecting letters to write arbitrary words.

Unigesture is a zoned input system. Examples of other zoned input systems include the number pad on a phone and Quikwriting [4]. Zoned input reduces the number of input choices from the number of letters in an alphabet to the number of zones offered by the zoned input system. Thus, zones can map letter selection to a mechanism with limited input options, such as a phone. Zoned input is an obvious choice for Unigesture because there are only so many ways to tilt one's hand while holding a device. Thus, we clump letters into zones, and assign each zone a reasonable tilt-based gesture. The figure below shows a possible zone layout for the Unigesture system.

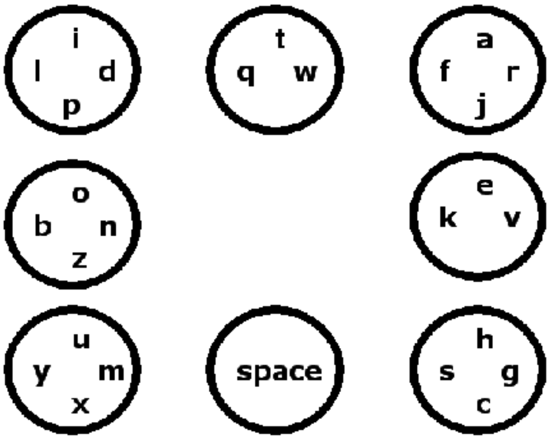

Figure 1. "Spread-Out" Letter Layout.

1.2 Letter Layout and Letter Inference

With Unigesture, the user selects a zone by tilting in the direction of a zone and then returning to his/her central or neutral position. Their are eight zones, seven of which include letters. The eighth zone is reserved for the most common English character, the space. As in all zoned input systems, there is the question of how to indicate which letter in the zone is wanted. Unigesture uses a method not unlike the T9 TM inference system [5] for cell phones. In Unigesture, the user never directly specifies which letter in the zone is wanted. Rather, the user enters an entire sequence of zones in order to form a word. Once a word has been entered, the system will attempt to infer which letters must have been wanted by the user. For example, with the letter map shown above, if the user tilts to the north zone, then to the southeast zone, and then to the east zone, followed by space, the system will infer that the user wants the word "the". 't' is located in the north zone, 'h' is in the southeast zone, and 'e' is in the east zone. No other combination of letters from that particular zone sequence makes sense as a common English word.

The letter layout shown above is called the "spread-out" letter layout because common letters are spread out among the zones. Each vowel is located in its own zone, and 't' and 'h' have their own zones as well. This design is intended to reduce the number of possible English words that could be formed from each zone sequence. For each pair of zones, there is only one likely pair of letters that would appear consecutively in an English word. Thus, when the system attempts to infer a word, it is more likely that it will return the correct one.

The "spread-out" letter layout described above attempts to reduce the number of mistakes that the inference system makes. The "clustered" letter layout, shown below, attempts to reduce the number of mistakes that the user makes. I had hypothesized that the cardinal zones (north, south, east and west) were easier to input than the diagonal zones (northeast, northwest, southeast and southwest). Thus, the "clustered" layout places common letters in these "easy" zones.

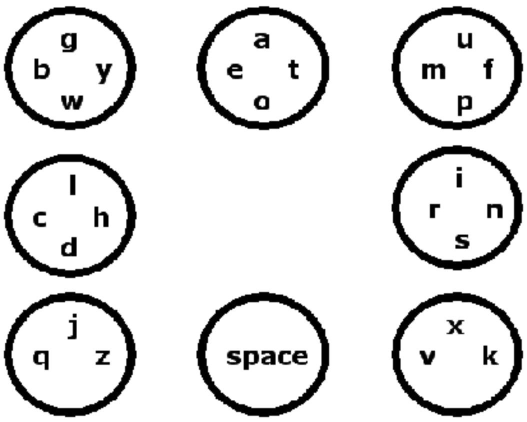

Figure 2. "Clustered" Letter Layout.

1.3 Degree of Tilt

Another variation in the Unigesture system is the degree of tilt. As stated above, the user selects a zone by tilting in the direction of a zone and then returning to his/her central position. There are two versions of of the tilt recognition engine – the "slight tilt" version and the "deep-tilt" version. The "slight-tilt" version requires a very short, fast tilt. The "deep-tilt" version requires a very long, deliberate tilt. The "deep-tilt" version places more strain on the wrist but requires less coordination than the "slight-tilt" version.

Figure 3. Tilting a small device. Pictured is a user tilting up and then returning to center.

2 Overview of the User Study

Plans for a user study were made as soon as a very preliminary version of Unigesture had been developed. However, the actual process proved to be somewhat bumpy. Originally, I had intended to use the study as a means to “evaluate” the Unigesture method. However, after talking to Judy Ramey, I learned that user studies are best used as design iteration tools than they are for formal evaluation. Moreover, I found myself perpetually postponing the user study, because I never felt my Unigesture prototype was “good enough” to unleash on users. When I changed focus and decided to use the user study to explore the design space of tilt-to-write methods, I found that even a rough prototype of Unigesture was sufficient. In addition, the knowledge learned from the study could help me improve the Unigesture implementation to the point where it was good enough to release to users.

2.1 Human Subjects Application

Initially I planned to conduct the user study at UW. My friend Kurt Partridge informed me that I may need to file paperwork with UW’s Human Subjects Division. After review of the Human Subjects Division’s webpage, I found that any study I would perform where users would actually use the Unigesture method would require full completion of Human Subjects paperwork. This is because all novel user interfaces are sufficiently different from common practice that any physiological effects need to be addressed. I found, however, that I could apply for “Expedited” review because I could argue that my study posed “Minimal Risk” to users.

Paperwork that must be submitted to the Human Subjects Division includes a formal application as well as any forms that you plan to give participants of your study. This includes consent forms, advertisements, and questionnaires. The application covers a number of topics. Part of the application asks open-ended questions, such as what the research protocol will be and what devices will be used. The remainder of the application lists concerns that the Human Subjects Division has and asks the applicant to state how he/she will address these concerns. Concerns include ensuring equitable subject representation, ensuring that recruitment will be done without loss of privacy or coercion, and ensuring risks will be disclosed and handled properly.

In the large part, I explained that this was a small study and that I was the sole person working on the study. Thus, I would personally ensure that most of the concerns were addressed. Some concerns, such as equitable subject representation across age groups and genders, I could not meet, due to the size of the study. However, I could help to ensure that users were not coerced by advertising only to groups and not individuals. I could also limit those participating in the study to those whose arms, wrists, and hands were healthy, thus limiting any users’ exposure to bodily harm.

The protocol for the experiment was expressed in a broad way in order to leave me the opportunity to change the study somewhat without losing Human Subjects Division approval. Each participant would undergo a training session, and after the training session would write words using Unigesture.

After filing this paperwork, I waited about a month for a response when I heard that the Human Subjects Division had gotten behind on responding to applications. When I called the Human Subjects Division, the coordinator of Minimal Risk Reviews promised me that she would do what she could but could make no promises.

As I was attempting to meet a paper deadline, I decided to move the user study from UW to Intel. Intel has a Center for Usability Engineering in Hillsboro, OR. Intel requires no application nor even a consent form. All they request is “screening criteria” for the participants. I requested participants who used either a cell phone or a PDA, because these participants would be familiar with technology held in the hand. The only new aspect for the participants would be the act of tilting the device as input.

Because Intel had no formal review process for studies involving user subjects, I could streamline my user study. I still had users fill out a consent form, but the consent form was much shorter than the one I had submitted to UW’s Human Subjects Division. I also removed steps such as rest periods between phrases in order to decrease the time needed for each user test.

2.2 Design of the Usability Study

The goals of the user study were determined jointly with my advisor Gaetano Borriello. AJ Bernheim Brush also helped me to take those goals and translate them into a valid experimental design. Gaetano and I had decided that the current Unigesture prototype was ready to act as a testbed for tilt-to-write design ideas. Thus, we designed the user study to explore the design space of what was possible. This exploration centered around two parameters – the letter layout used and the degree of tilt. As described above, there were two letter layouts as well as two levels of tilt , making a total of 4 varieties of Unigesture.

The prototype version of Unigesture used in the user study ran on a desktop PC. The user held a block in their hand which represented a handheld computer. The block (actually a deck of cards) had an accelerometer evaluation board attached to it. The evaluation board converted analog accelerometer readings into an RS-232 serial format, and this data was sent via a cable to a PC. On the PC, a small GUI presented user feedback as well as a reference letter layout map for the user.

12 users, all Intel employees, participated in this 4-way study. These participants were recruited by Intel’s Center for Usability Engineering. Each user was given either the “spread-out” or the “clustered” layout. The user would use only one letter layout for the duration of the test. The user was trained, first by watching me use Unigesture, followed by a time period where they were allowed to use Unigesture until they felt comfortable. The users were each allowed a total of 20 minutes training time; however, they were not told that explicitly, and instead stopped whenever they felt ready, which was usually after far fewer than 20 minutes had elapsed. Training was focused less on reaching mastery with the method and more on ensuring that the user had no misconceptions as to how Unigesture worked.

The user was then asked to write 8 phrases. 4 were written using the “slight-tilt” version and 4 phrases were written using the “deep-tilt” version. The entire experiment was counterbalanced.

Several pilot tests were conducted before the actual study went underway. These pilots helped me estimate how much I could realistically record in real time on my data entry sheets. I also used the pilot tests to refine my “script” for training. In addition, the pilot tests also uncovered some bugs in the prototype which had gone undetected. I did two pilot tests at UW and one at Intel.

The order of events in each user test were as follows:

1. Welcome / Consent Form
2. User Fills Out Questionnaire
3. User Trains
4. User Enters Four Phrases
5. New Version Started
6. User Enters Four Phrases
7. Follow-Up

The questionnaire asked basic questions about past computer and cell phone experience. The questionnaire has been appended to this report. Data collected during the experiment included the time to train, the time to complete each phrase, and the number of mistakes made. Audio was also recorded as the user entered phrases and answered follow-up questions.

An example of a data collection sheet has been appended to this report. A checkmark means that the user correctly entered the appropriate zone, while an ‘X’ means that the user made a mistake.

Follow-Up questions included the following:

1. Did you prefer the “Deep-Tilt” version or the “Slight-Tilt” version?
2. Which zones were easy/hard?
3. Did you feel any physical discomfort?

3 Results of the User Study

For full results of the user study, I refer you to [1]. In this document, I will only state the most important results of the study very briefly.

We were looking for potential effects that the 4 different versions of Unigesture had on user performance. For this part of the study, we measured performance in terms of number of mistakes made per phrase entered. We found that user performance was very variable, and that the natural variation among users affected performance much more than any aspect of the experimental setup (“spread-out” versus “clustered”, and “slight-tilt” over “deep-tilt”). The reasons for this variability included differences in how people held the mockup device, whether the user was right or left-handed, and whether the user played video games. Most users who mastered the “slight-tilt” version preferred it, but those users who had great trouble with the “slight-tilt” version could adequately use the “deep-tilt” version.

Each user had their own unique set of “problem zones”. This resulted in lower-than-expected performance for the “clustered” letter layout. Please recall that the “clustered” letter layout had clumped common letters into the cardinal zones because it was hypothesized that the diagonal zones were harder to reach. Diagonal zones were commonly troublesome, but not always, and thus the letter layout was not actually as helpful as we hypothesized it would be. However, I believe that the concept of clustering common letters in easy zones is still valid. What Unigesture needs is an automated training program which can evaluate which zones are easy for the user and then place the common letters in those zones.

Of the 12 users, 5 said that they experienced no physical discomfort, 4 of the 12 experienced slight physical discomfort, 2 users felt fatigued, and 1 user felt pain. Qualitative responses to the system were largely positive. Only 1 of the 12 users said something negative about the system; this user found Unigesture “annoying”.

Our usability study addressed a number of other questions, such as whether learning occurred during the course of the user test. We also did a mini-study that attempted to study how much time novice users spent searching for letters on an unfamiliar letter layout map. Details of these results are in [1].

The conclusion of our user study was that a semi-personalized system would be best. As stated above, the goal would be to determine the “problem zones” for each user and then cluster letters in zones that are not troublesome. Another option is to increase the granularity of Unigesture and map common words or phrases to easy zones instead of letters. However, it is not clear yet whether the 8-zone approach of Unigesture is the best approach for whole-word entry. Tilting to scroll through a list of whole words may be a better option, because tilt-to-scroll requires no training to master.

4 Commentary on the User Study

The Unigesture user study was a great learning experience for me. Throughout the process I was confronted by decisions, tasks and situations that I do not usually experience as a software engineering researcher. While my observations will hardly startle the experienced usability tester, I believe that I learned enough to make my next user study considerably smoother.

4.1 Script or No Script?

One of my first concerns was the script of the usability study. At first, I dutifully wrote down a literal script of how I would train users. The script was long, and I found that inevitably I would lose my place in the script and suffer an awkward pause when I tried to find my place. I also found that while I remembered my “lines” quite easily, I often forgot when to take actions (turn on tape recorder, pick up stopwatch, etc.) So I replaced my word-by-word script with a “test plan” that sketched out brief summaries of what I would say, separated by actions that I needed to perform. I could easily add items to the “test plan” and I found I was able to successfully remember all the parts of the user study by referring to it. The "test plan" was also a manageable size.

4.2 Pilot Tests and Mistake Management

The second observation I had involved pilot testing. As stated earlier in the paper, I found the pilot tests enormously helpful; I improved my test plan, identified bugs, and streamlined the way I recorded data. However, I think the most important effect of pilot testing was reducing my nervousness. I was incredibly nervous. Some of this nervousness was quite justified – unlike writing code, where you can always backspace, I had to be right on target without making mistakes during every moment of each user test. Otherwise, the user test would have to be discarded. Also, since there was some quantitative data being measured, I had to be accurate as I timed users and recorded errors or I would risk reporting bad data. The pilot tests convinced me that this task was doable.

When I did make a mistake, I found that the audio recording of the user test was quite useful. Although I'd made no deliberate audio cues during each test, there was enough context spoken by myself and the user to effectively time users from the audio recording if necessary. Luckily I only needed to retrieve data in this manner once.

4.3 Wait, Let Me Take Off My Developer Hat and Put On My Usability Hat

I initially found it incredibly difficult to detach myself from the Unigesture software that I had written. During the pilot tests, I would see the system do something strange, wonder why the system had done that, start evaluating possible causes in my brain, and in the meantime, neglect to watch the user! AJ Bernheim Brush had warned me of this "developer's distraction", and she couldn't have been more correct! People do things with your software that you never dreamed people would do. At first, this was a shock to me. By the time the formal user study began, I had learned to carefully note all the startling things I saw but not to ponder on it until after the user had left.

I also found it was very easy to get depressed during the user study. At least one participant asked me if I had designed the accelerometer hardware, and I explained that I hadn't. I had purchased the accelerometer evaluation board. He then asked me what work I *had* done. I, of course, had written all the software and had also designed the user study. I could surmise that some Intel hardware designers did not consider my contributions to be "real work"!

More seriously, I found it was quite difficult to stay upbeat and objective when I saw a few users struggle to learn Unigesture. I was personally responsible for their difficulties. Perhaps this is a good experience for a developer to go through, but I think it may be superior to have an unattached usability specialist give the test and have the developer experience this sorrow while watching a videotape (grin).

5 Final Thoughts

This document narrates my experiences conducting a usability study of the Unigesture text entry method. The results of the study were interesting, although not necessarily the results for which I had hoped. Nonetheless, the process itself was a great learning experience. For example, I received a taste of the rigor required of medical and social science experiments. I also gained an appreciation of the immense effort required to conduct a usability study and examine its results. Finally, I learned how time spent early in the process on user study design and pilot testing can ensure a successful user study later on in the process. I believe that these lessons will stay with me long after the Unigesture project finishes.

6 Postscript

My application to the Human Subjects Division was approved on 9 May, 2001, after a change to the Consent Form was requested. I completed the user study at Intel on 3 May, 2001.

References

[1] Vibha Sazawal, Gaetano Borriello, and Roy Want. Experiences with unigesture: A one-handed text entry method for small devices. Submitted to ACM UIST 2001.

[2] Kenneth Fishkin, Anuj Gujar, Beverly Harrison, and Roy Want. Embodied user interfaces for really direct manipulation. *Communications of the ACM*, 43(9), September 2000.

[3] William Hamburgen, Deborah Wallach, Marc Viredaz, Lawrence Brakmo, Carl Waldspurger, Joel Bartlett, Timothy Mann, and Keith Frakas. Itsy: Stretching the bounds of mobile computing. *IEEE Computer*, April 2001.

[4] Ken Perlin. Quikwriting: Continuous stylus-based text entry. In *Proceedings of ACM UIST '98*, 1998.

[5] Tegic Communications. T9 web page. URL: http://www.t9.com.