COMPARISON OF SPEECH WITH KEYBOARD AND MOUSE AS THE TEXT ENTRY METHOD

Dr. Sehchang Hah and Vicki Ahlstrom
FAA, William J. Hughes Technical Center
Atlantic City International Airport, New Jersey

Federal Aviation Administration (FAA) researchers ran an experiment that compared an automatic speech recognition system with the keyboard and mouse as text input methods. In the speech condition, the participants read the document in its entirety, and the system converted speech into text. They corrected errors using the keyboard and mouse. In the typing condition, the participants corrected errors as they typed. The experimental results showed they spent significantly less time reading than typing to enter the text. When we factored in correction time for both conditions, the results showed that participants took significantly more time in the speech than in the typing condition. Whether they were fast or slow typists, all of the participants preferred typing to speech and performed better in the typing condition than in the speech condition. Optimization of the system and more training may improve the performance of the speech recognition system.

INTRODUCTION

Speech is the natural communication method for humans. With technological advancement, we now use various methods to communicate. Reports are normally typed. However, some people do not type well. An efficient way to produce reports could be to convert speech into documents using speech-recognition systems. FAA Operational Control Center (OCC) specialists regularly type to report National Airspace System (NAS) equipment and systems delays. This research evaluated a continuous speech recognition system to determine if it was a more efficient method than the keyboard and mouse for entering accurate technical information.

For continuous-speech applications, users train the system for their speech patterns first. As users speak to a speech-recognition system, the system extracts their speech every 10 to 30 milliseconds into sound elements utilizing speech characteristics already collected during training. The recognition system uses these elements to look for the most probable word based on acoustic, lexical, and language models (Zue & Cole, 1997).

In 1997, researchers (Mogford, Rosiles, Wagner, & Allendoerfer) at the William J. Hughes Technical Center, Research Development and Human Factors Laboratory (RDHFL) conducted a study to assess the maturity of speech recognition systems for use by technical specialists. The participants performed an antenna-alignment procedure. In the speech condition, the specialists first trained a speech recognition system with 19 command words and then used the words in addition to seven system words. In the traditional method condition, they used hard-copy manuals. Overall, they found that the technology was not mature enough for use in an operational environment. Given the technological advancement of speech recognition systems in recent years, we decided to compare one type with the keyboard-mouse input for continuous speech applications.

We were also interested in the relationship between a person's typing skill and performance in both the typing and speech conditions. Mitchard & Winkles (2002) reported speech was better than keyboard and mouse as the input method for users who typed less than 45 words per minute. However, the keyboard and mouse method was better for those who typed more than 45 words per minute.

METHOD

Participants

Six local males and six females volunteered to participate in the current experiment that was run in an office setting. We randomly assigned them to experimental conditions. All of the participants spoke English as their first language.

Apparatus

We used Dragon NaturallySpeaking 7.0 by ScanSoft (2003) for the speech condition. It converted the participants' speech into text files. We installed the software on a Dell Precision 340 that had a Pentium 4 processor with 2.0 GHz speed with 514 KB Level 2 (L2) Cache, 100 MHz bus speed, and 512 MB Error-Correcting Code (ECC) Rambus Direct Random Access

Memory (RDRAM). The computer used Soundmax Integrated Digital Audio by Analog Devices, Inc. The participants spoke into a microphone of an antinoise Platronics DSP-400 headset. The computer capability exceeded the requirements recommended by ScanSoft. The keyboard was a QWERTY-type Gateway Model-G9900, and the mouse was a three-button Dell IntelliMouse 1.3A.

Our in-house software created an interface between Dragon NaturallySpeaking and Windows. It supplied the experimental data collection routines. It recorded the time and content of participants' utterances, typing inputs, and speech-recognition system responses.

After the data collection trials, we administered a questionnaire to collect the participants' opinions on the two methods. To examine the relationship between their typing skills and performance in the experiment, we tested their typing skills with TypingMaster Pro 2002.

Material

The participants used two documents: Document 1 had 66 words and Document 2 had 69 words. They used the same documents for both conditions. The documents were the remarks that OCC specialists had entered on the NAS Equipment and Related Delays notice at http://technet.faa.gov. In the keyboard and mouse condition, they used Microsoft WordPad to type them. In the speech condition, they read them to the speech recognition system. The participants verbally conveyed special characters such as "open parenthesis" for "(," which made them speak 159 words for Document 1 and 155 words for Document 2, respectively.

Experimental design

We used two input methods and two documents as experimental variables. We used a randomized block factorial design, and each participant received all combinations of the experimental variables.

Procedure

Before the experiment, the participants read a document for about five minutes to train the speech recognition system. After training the system, the participants practiced using a practice document. During the practice trial of the speech condition, the experimenter explained specific ways to speak acronyms and special characters such as "open parenthesis" for "(." They did not have to memorize how to speak special characters because those were written and embedded in the document.

After the practice, the participants performed data collection trials. They were asked to read and type as quickly and accurately as possible. The experiment lasted about an hour. Spell-checker was not available for their use for either condition of the experiment.

In the speech condition, the participants read aloud a document placed at their eye-level. After they read the whole document, they pressed the F2 key. This made the document disappear from the display, and a prompt appeared. As soon as the participants responded to the prompt, the error-correction session started. The document was displayed in Microsoft WordPad. They corrected errors with the keyboard and mouse because error correction with speech has been inefficient (Halverson, Horn, Karat, & Karat, 1999; Eu & Hedge, 1999; Mitchard & Winkles, 2002). After they finished correcting the document, they pressed the F2 key to end the trial.

In the keyboard and mouse condition, the participants read and typed the same document in Microsoft WordPad. The participants were asked to correct errors as they typed. After they finished typing the document, they pressed the F2 key to end the trial.

We measured the time to perform the task and errors. After the data collection trials, they completed a questionnaire and took a 5-minute typing test of TypingMaster Pro 2002.

RESULTS

Time Measure

There was no significant time difference between the two documents for both typing (Sign test, n = 12, $T_+ = 9$, $p > .05$) and speech conditions (Sign test, n = 12, $T_+ = 8$, $p > .05$).

The participants read the documents significantly faster without corrections (1 minute and 15 seconds, *sd* = 14 seconds) than typing with corrections (2 minutes and 34 seconds, *sd* = 48 seconds) [Sign test (n = 24, $T_+ = 22$, $p < .01$)].

When we factored in the correction time for both conditions, they took significantly longer in the speech condition (3 minutes and 52 seconds, *sd* = 63 seconds) than in the typing condition (2 minutes and 34 seconds, *sd* = 48 seconds) [Sign test (n = 24, $T_+ = 2$, $p < .01$)].

The range of participants' speech rates was from 81 to 141 words per minute for the 159-word Document 1 and from 82 to 142 words per minute for the 155-word Document 2. The average rates were 114 words per minute for Document 1 and 111 words per minute for Document 2, respectively. These words included special-character words such as "open parenthesis" that were counted as two words.

Even the participants whose typing speed was slow performed better with the keyboard and mouse than with speech as shown in Figure 1. In the figure, we present the average data from both documents.

The participants' time to complete entering the documents using the speech recognition system and the keyboard and mouse showed high correlations with their typing speed measured by TypingMaster Pro 2002. The Speaman's rhos between the speech and the typing speed and between the keyboard and mouse and the typing speed were -.60 and -.86, respectively. This must be due to the fact that in both the speech and the keyboard and mouse conditions, the participants corrected errors using the keyboard and mouse.

Errors

The speech recognition system produced many substitution and insertion errors, that is, words. For the 159-word Document 1, there were 21 substitution errors (*sd* = 8), 12 insertion errors (*sd* = 10), and 3 omission errors (*sd* = 1). For the 155-word Document 2, there were 17 substitution errors (*sd* = 9), 10 insertion errors (*sd* = 8), and 1 omission error (*sd* = 2). The finished documents after error corrections did not have many errors. For the typing condition, the mean for total errors was 3 words (2.2%) for both documents and for the speech condition, it was 6 words (4.2%).

Questionnaire responses

All participants, except two, rated speech more difficult to use than the keyboard and mouse. The two participants rated speech easier to use than the keyboard and mouse. All participants had to speak special character commands such as "caps on," which were not familiar to them. They mentioned that the speech-recognition system errors were hard to find and predict. All preferred typing.

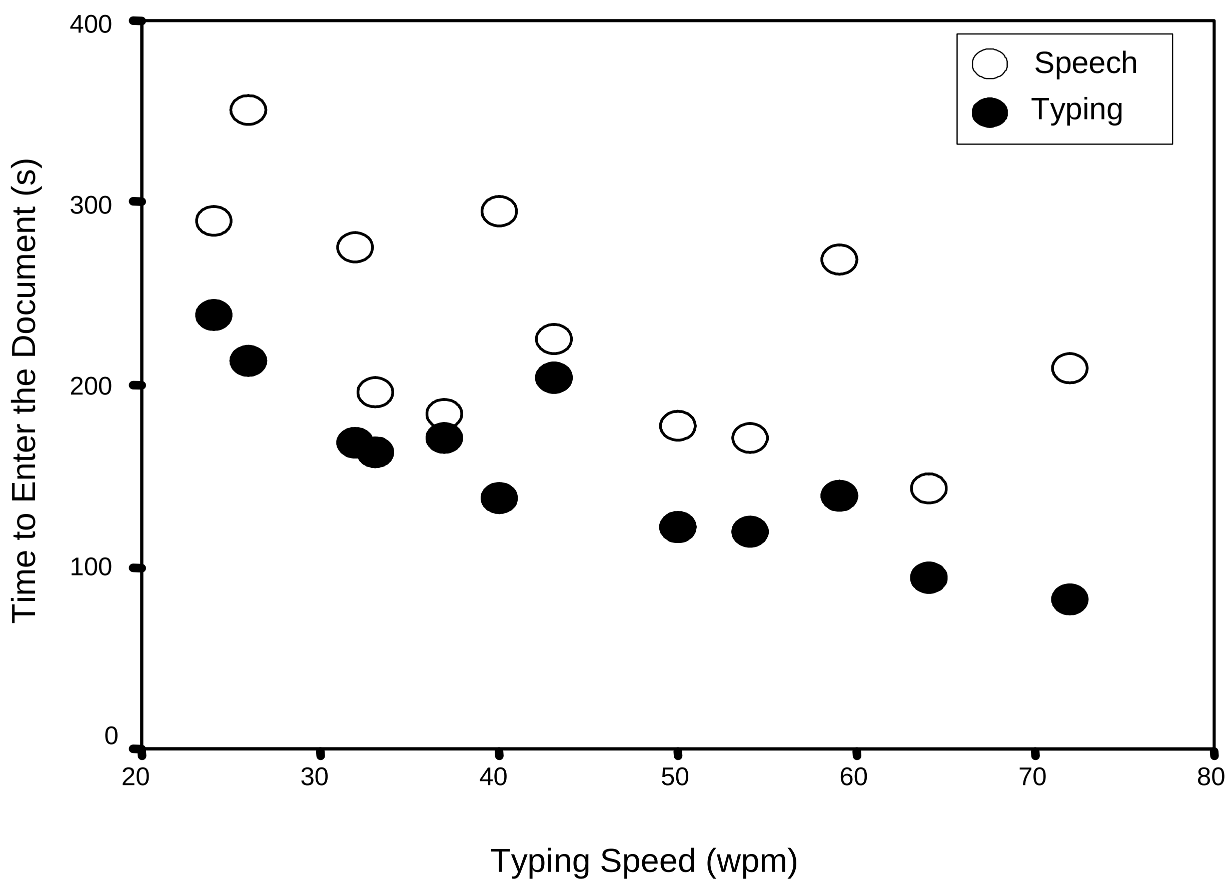

Figure 1. Participants' average text entry time including correction times of speech and keyboard and mouse conditions to their typing speed measured by TypingMaster Pro 2002.

DISCUSSION

When we factored in the correction time for both conditions, the speech condition turned out to be less efficient. All participants preferred typing to speech. The participants mentioned that in the typing condition, as they made a mistake, they noticed and corrected it. For the speech condition, because they read aloud the document in its entirety first and corrected errors later with keyboard and mouse, they needed to verify the speech recognition errors with the original document before correcting the errors. Because the speech recognition system's errors were incorrect words, but correct in spelling, the participants needed to check them on the phrase- or sentence level beyond the word-level.

It appears that the keyboard and mouse was superior and the participants preferred it. Speech recognition was not mature enough for this application given our sample of computer-experienced participants.

Mitchard & Winkles (2002) reported that slow typists entered phrases better with speech than with keyboard and mouse and that fast typists did better with keyboard and mouse than with speech.

However, our results showed no difference as a function of typing speed (Figure 1). Our results may be due in part to the difference between their task and ours. Their participants read a phrase such as "NEW COURSE REQUIRED," and our participants read a paragraph. In addition, our documents had many special characters while theirs did not.

If our speech recognition system committed fewer recognition errors, the participants might have finished the task faster with speech. To alleviate this problem, we can optimize the system and provide more training. One of the optimization methods is to replace acronyms with simpler words. For instance, we could ask participants to speak "arts" for "Automated Radar Terminal System (ARTS)" instead of "cap automated cap radar cap terminal cap system open parenthesis caps on a r t s caps off close parenthesis." Such methods will reduce the number of words to speak and thus the number of errors. It will also make reading the documents easier.

ACKNOWLEDGEMENTS

The authors wish to thank Gary Mueller and Yev Tabekman, Titan Corporation, for their computer programming support, Dr. Earl Stein and Dr. Mike McAnulty, ATO-P, for their review of the paper, Denisse Villa, Precision Infrastructure, Inc., for her secretarial support and data summary, and Linda Johnson, Northrop Grumman Information Technology, for her editing support. In addition, they thank Helen Mitchard, David Williamson, Timothy Barry, and Gregory Barbato for their helpful information for the research. The authors would like to especially express their gratitude to their colleagues who volunteered to participate in the experiment.

REFERENCES

Eu, H. & Hedge, A. (1999). Survey of continuous speech recognition software usability. Ithaca, NY: Cornell University. Retrieved April 5, 2004, from http://ergo.human.cornell.edu/AHProjects/Hsin99/Voice%20Recognition%Paper.pdf

Halverson, C. A., Horn, D. B., Karat, C, & Karat, J. (1999). The beauty of errors: Patterns of error correction in desktop speech system. *Proceedings of the Interact '99 Seventh IFIP Conference on Human-Computer interaction, Incorporating HCI '99* (pp. 133-140). Edinburgh, Scotland: IOS Press. Retrieved April 5, 2004, from www.research/ibm.com/SocialComputing/ChristineHalverson.htm

Mitchard, H. & Winkles, J. (2002). Experimental comparisons of data entry by automated speech recognition, keyboard, and mouse. *Human Factors, 44(2)*, 198-209.

Mogford, R., Rosiles, A., Wagner, D., & Allendoerfer, K. (1997). *Voice technology study report (DOT/FAA/CT-TN97/2)*. Atlantic City International Airport, NJ: FAA William J. Hughes Technical Center.

ScanSoft. (2003). ScanSoft - Dragon NaturallySpeaking 7.0. Retrieved April 14, 2003, from http://www.dragonsys.com/naturallyspeaking/

Zue, V., & Cole, R. (1997). Chapter 1. Spoken language input. In G. B. Varile & A. Zampolli (Eds.), *Survey of the state of the art in human language* (pp. 1-49). New York, NY: Cambridge University Press.