Use of Speech Technology in Real Life Environment

Ruimin Hu[1], Shaojian Zhu[2], Jinjuan Feng[1], and Andrew Sears[2]

[1] Computer and Information Sciences Department, Towson University, Towson, MD 21252
{jfeng,rhu1}@towson.edu
[2] Department of Information Systems, Interactive Systems Research Center UMBC, Baltimore, MD 21250
{szhu1,asears}@umbc.edu

Abstract. This paper reports results from two six-month field studies investigating the use of speech-based interactions in real world environments. The first study focused on the use of speech-based dictation/word processing tasks that allow users to generate text such as reports or emails. The second study took a broader view and examined two types of interactions: speech-based dictation for word processing and speech-based command and control supporting interactions with the graphical user interface. The results suggest that user requirements for speech-based interactions have evolved as the technology matured to include better support for formatting text documents as well as more effective support for other applications. While significant research has focused on the use of speech recognition to generate text, our participants spent substantially more time using speech recognition for other, often unexpected tasks such as instant messaging and web browsing. The use of speech recognition to support informal communication is particularly interesting as accuracy may be less critical in this context. Further research is needed to address these emerging requirements for speech technologies.

Keywords: Speech-based Application, Speech Interaction, Speech Technology, Physical Impairment, Field Study

1 Introduction

Speech-based applications allow users to communicate with computers or computer-related devices without the use of traditional input devices such as the keyboard and mouse. As a result, speech technologies are of particular importance for some individuals with physical disabilities. While multiple studies have investigated the use of speech-based input by individuals with physical disabilities, most were lab-based studies that focused on pre-defined tasks that were somewhat, but not completely, representative of the everyday tasks and environments [20]. In addition, these studies have typically been quite short in duration, ranging from one to ten hours [6]. As a result, we have limited knowledge of how people may use speech recognition as they

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 62–71, 2011.

interact with computers as part of their daily activities in realistic environments over an extended period of time.

To begin filling this gap, we conducted two six-month field studies investigating how individuals with and without physical disabilities use speech technologies when interacting with desktop computers in their home environment. The first study focused on the use of speech recognition for dictation/word processing tasks such as writing reports or emails. The second study took a broader view and examined two types of interactions: speech-based dictation for word processing and speech-based command and control to support interactions with the graphical user interface.

2 Related Research

Automatic Speech Recognition (ASR) has experienced significant commercial success in recent decades [13]. Traditional dictation software such as Dragon NatuallySpeaking offers an improved user interface and higher recognition accuracy than in the past. The integration of speech recognition engines into commercial operating systems makes ASR readily available for the general public. Simultaneously, potential applications of ASR have become more numerous, covering a variety of tasks beyond the traditional domain of text documents. For example, both Dragon NatuallySpeaking and built-in speech recognition support in Windows Vista™ both support web browsing and desktop navigation. With all of these recent developments, there is limited research examining if and how these new functions are being used.

Speech-based dictation systems have been the traditional focus of both academic research and industry. For example, numerous hands-free ASR error correction techniques have been explored (e.g. [2]). Halverson et al. [8] studied error correction in large vocabulary, continuous speech recognition systems and identified different error correction patterns depending on the users' experience with speech-based interactions. A number of studies have investigated speech-based navigation within the context of text documents. Among them, McNair and Waibel [17] investigated an early version of target-based navigation where incorrect words were selected via speech. Sears et al. [20, 21] investigated how individuals with high level spinal cord injuries (SCI) employed speech recognition systems to compose text documents. No differences were identified between participants with and without SCI in terms of recognition error rates, navigation command failure rates, or overall productivity. However, participants with SCI interrupted their dictation more frequently to correct errors and were more satisfied with their interactions. In addition, two long-term studies have investigated the use of speech-based dictation systems by individuals with motor and cognitive disabilities, but both studies included only one participant making it difficult to generalize the findings [16][19].

Speech-based command and control can support interactions with typical graphical user interfaces rather than text documents. Three techniques are currently available: target-based selection, speech-controlled cursors, and grid-based selection. Manaris and Harkreader [15] explored on the use of speech recognition to generate keystrokes and mouse events. Karimullah and Sears [14] focused explicitly on cursor control, evaluating the efficacy of a predictive cursor designed to help users compensate for processing delays associated with speech recognition. Many other approaches for

controlling a cursor have been explored. For example, Mihara et al. [18] discussed a system in which multiple 'ghost' cursors are aligned vertically or horizontally with the actual cursor while de Mauro et al. [5] investigated a voice-controlled mouse that used individual vowels as commands. Harada et al. [9] explored a similar idea called a vocal joystick that allowed users to control the cursor by varying vocal parameters, i.e., vowel quality, loudness, and pitch. Finally, grid-based solutions position the cursor using recursive grids allowing users to 'drill down' until the cursor is in the desired location [12] with two variations of the grid-based solution being compared by Dai et al. [4].

A few studies examined speech-based interactions in different domains. For example, Christian et al. [3] investigated speech-based navigation in the context of the web while Arnold et al. [1] examined speech-based programming system that integrated speech recognition and a predefined syntax for the programming language. Similarly, Hubbell et al. [11] developed a syntax-directed graphical editor for programmers with physical impairments. Sporka et al. [22] explored the control of computer games using speech and non-verbal vocalizations while Harada et al. [10] developed Voicedraw, a completely hands-free speech application for generating free-form drawings. Both speech and non-speech voice-based interaction techniques were adopted.

These studies provided insights into the effective use of speech-based interactions. Clearly, speech-based solutions are more appropriate for certain types of tasks and users. An important limitation of most previous studies is that the evaluations were conducted in less than realistic environments with artificial tasks. While several studies have evaluated speech-based interaction in more realistic contexts, such as web navigation or gaming, the focus was on a single task, isolating users and ensuring that they would not interact with rest of the computing environment. Given the variety of tasks users perform on computers, and the variety of speech-based interactions that can be supported on personal computers, there are important questions to be answered regarding how users will interact with speech applications in realistic settings. In the current field studies, we investigate how users with and without physical disabilities use speech to interact with computers in their everyday lives.

3 First Field Study

3.1 Methodology

The first study focused on the use speech-based dictation to generate text documents. Five participants took part in this study (see table 1). All participants used computers provided by the researchers with exactly the same hardware and software specifications. Computers were provided for two reasons. First, some participants did not have their own computers at the time of the study. Second, some participants who had computers used different operating systems and applications. It is difficult to provide users with comparable speech-based interaction experiences if they are using different hardware or operating systems; thus it would be difficult to compare results among users. Providing computers for all participants addressed these problems. Several standard Windows Vista speech functions were modified for use in this study

including one method of navigating within dictated text and a second technique that supported navigation within any application.

Dictation and Editing. The Vista speech system offers two ways of navigating within text documents. Target-based navigation allows users to select words by stating the navigation target (e.g., 'select book'). Direction-based navigation allows users to select targets by specifying the movement direction and units (e.g., 'move up five lines'). We added anchor-based navigation to provide more flexibility in error correction [7].

Desktop and Menu Interaction. Vista offers a grid-based navigation mechanism. Currently there is no limit to how many times a user can zoom in to a smaller portion of the screen, but after three or four levels it becomes difficult to identify either the target or grid numbers. We modified the existing grid-based solution to disable zooming after three levels at which point a magnification function was enabled to enlarge the selected grid. At this point, four simple commands ('Up', 'Down', 'Right' and 'Left') were enabled to allow users to fine-tune the cursor location. For detailed information on the effect of the magnification function see [23].

Table 1. Age and description of disabilities

Users	Age	Physical impairments
S1-P1	60	Arthritis in neck, spine, hands and wrists (carpal tunnel)
S1-P2	19	Duchene muscular dystrophy
S1-P3	34	Dexterity with M.S. and left hand
S1-P4	48	Stroke; Arthritis limits upper arm, hand & neck movements
S1-P5	41	Significant weakness in hands & arm

The study was conducted at the participants' homes and lasted for six months. During the first visit, the researchers demonstrated the speech functions offered in Microsoft Vista including the two modifications outlined above. Participants were instructed to use the computer and speech applications for whatever tasks they need to complete, but they were required to generate at least four pages of text with no restrictions on the content. Interviews were conducted at the end of each month to collect feedback and logged data.

3.2 Results

Overall, all participants increased how quickly they generated text in the first several months, indicating the positive learning effect. However, all participants also slowed down after the initial improvement and there were notable fluctuations in performance. For three participants (P1, P2, and P4), the average number of words generated is lower compared to results from earlier lab-based studies.

Quality of text generated the study was assessed based on the percentage of sentences containing errors (including both recognition and grammatical errors). The text documents generated by participants did not show a consistent trend with regard to

quality. Fluctuation was very common. The quality of the documents produced by P1 and P2 improved during the observation period, but there was no clear trend for P3, P4, and P5.

Efficiency (WPM) does not appear to be a decisive factor for user satisfaction and future adoption of the technology. P1 and P4 had similar entry speed (3.7 vs. 5.1) with contrasting satisfaction and attitude. P1 was very negative and would not continue using the technology while P4 was very positive throughout the study. P5 was pretty fast (13.3) but held a neutral attitude towards adoption while both P2 and P3 (8.7 and 11.8 respectively) held a positive attitude towards adoption.

4 Second Field Study

4.1 Methodology

The second study focused on a broader context and investigated both speech-based dictation for word processing and speech-based command and control to support interactions with the graphical user interface. Ten participants took part in this study, five had no physical disabilities and five had disabilities (see table 2). The study employed the same procedure as the first field study except that the participants were instructed to use the computer and speech applications for whatever tasks they would like to complete.

Table 2. Age and description of disabilities

Users	Age	Physical impairments
S2-P1	57	Severe Carpal Tunnel Syndrome
S2-P2	57	Injury resulting in muscle weakness and lack of sensation in hands and arms
S2-P3	37	High level Spinal Cord Injury (C5, 6)
S2-P4	38	Stroke
S2-P5	58	Multiple Sclerosis

4.2 Results - Differences between Two Groups

Applications Used/Tasks Completed. The participants without disabilities used a larger variety of applications than those with disabilities. Although both groups used the speech functions in Word, Internet, Outlook, IM, and Desktop navigation, only Word was used by every participant. In addition to these applications, the participants without disabilities also used speech commands in the context of PowerPoint, programming, games, and Notepad. The users with disabilities focused more on using Word and Internet Explorer, which served their basic needs. Table 3 summarizes the applications used by the participants in both groups.

Table 3. Applications used by both groups

Applications Used	Users without Disabilities	Users with Disabilities
Word	X	X
Internet/email	X	X
Outlook/email	X	X
Desktop navigation	X	X
IM	X	X
Music and videos		X
PowerPoint	X	
Programming	X	
Game	X	
Notepad	X	

Frequency of Use. To our surprise, the participants without disabilities used the speech functions more frequently than those with disabilities. We observed that the infrequent usage of speech functions by the participants with disabilities seemed to be associated with a general lack of interest in computer usage rather than the speech functions themselves. Two of these participants did not have a computer at the time that the study started, so a computer was not an integral component of their daily lives. In general, these participants experienced difficulty generating text documents, so they had developed strategies to avoid those tasks.

Evolving Requirements for Speech Applications. This study confirmed a number of well-known challenges regarding speech-based applications, such as frequent recognition errors and inefficient navigation. More interestingly, the results revealed a number of emerging requirements that may serve as the focus for future research.

Effective Editing and Formatting Functions. In the past, recognition errors were widely acknowledged as a major problem for speech-based dictation applications. Consistent with these concerns, participants in our earlier lab studies used editing functions to fix recognition errors but they did not spend time worrying about the appearance of the resulting text. In contrast, the participants in the field study were very interested in fixing the appearance of the generated text documents. They spent substantial time trying to make the document 'look right'. They cared about the details that participants in our lab studies would typically ignore. For example, participants frequently made changes to punctuation, fonts, and how text was aligned. During interviews, participants expressed a desire to learn more and to become proficient using the editing and formatting commands.

Most speech-based dictation systems offer editing and formatting functions. However, due to the large number of editing and formatting functions available, most commands are hidden in a lower level of the menu and are not directly visible on the screen. As a result, individuals needed to have substantial knowledge about the Word

menu bar and where specific editing or formatting functions could be found. This created more problems for the participants with disabilities because they had not used Word as often. Two additional challenges created problems with editing and formatting: commands and dictation are still frequently confused and it is still too difficult to select text that needs to be formatted.

Integrated Desktop Interaction. Previous studies tended to examine speech-based interactions in a single context, such as text generation, drawing, or desktop navigation. During the field study, participants used speech in multiple applications and contexts and frequently switched between applications (see table 3). This pattern suggests that providing a consistent speech-based solution, which can be used regardless of the application, is critical. Consistent dialogue design among applications would also improve usability. Finally, participants expressed a need for additional functions, such as the ability to use speech to emulate short-cut keys.

Web Browsing. All participants with disabilities expressed great interest in using speech commands to access information on the Web, including those who had limited computing experience. These participants spent substantially more time browsing the Web than they did using Word and Outlook combined. Currently, the Vista speech environment offers limited web browsing functions, allowing users to move between links or to say the text associated with a link to open a target page. Entering a URL is a significant problem for the participants. Some participants used the keyboard to enter URLs, even though this was a rather slow process while others accessed web sites using the 'favorites' function or desktop icons.

Online Communication. Interestingly, multiple participants without disabilities and one with a disability used speech for online communication activities via IM. They commented that it was faster and that they experienced fewer problems regarding editing and formatting. They were more tolerant of recognition errors in IM than when using Word or Outlook. The participant with a disability who used speech for IM specifically commented that he would be reluctant to adopt speech for writing text documents but preferred using it for instant messaging because it was faster.

Data Entry in non-text Environment. Participants also expressed the need to use speech to enter data in non-text environments such as an online calendar or a spreadsheet. Currently, speech interactions are rather difficult under both circumstances. In Excel, users cannot directly dictate a word or a number. There are no effective methods for selecting specific columns or rows for formatting or analysis purposes. Participants also experienced problems positioning the cursor in a specific cell because the cursor frequently jumped to items in the menu bar. Calendar applications typically allow users to dictate event descriptions, but navigation was problematic. Data entry in a calendar requires the user to select specific time slots, which was difficult using speech.

Entertainment Applications. Multiple participants used speech to play games. Participants without disabilities were more interested in battle games with rich graphics and sound effects. They adopted multi-modal interaction strategies such as using the mouse to specify targets and speech to 'double click'. Participants with disabilities did not use games during the field study, but they expressed interests in using speech to play games such as poker or puzzles. One participant with disability did use speech to manage the songs on his iPhone and computer.

5 Discussion

The most noteworthy contribution of these field studies is highlighting the variety of applications used by participants and, as a consequence, newly emerging requirements for speech-based interactions. The findings confirm that interest in speech-based interactions extend beyond generating text. Within the traditional domain of text generation, participants demanded more powerful and comprehensive editing and formatting capabilities that not only allow them to correct errors, but also allow them to effectively manipulate the appearance of the document. This represents a very different scenario as compared to what has served as the focus of most related research. At the same time, this is a challenging task given the capabilities available in most systems.

Interestingly, a few participants (from both studies) developed similar multi-modal interaction strategies. Participants with disabilities leveraged the residual control they had of their hands and arms. These participants could control the mouse movement to some extent but clicking and double clicking were very difficult, so they would use the mouse to position the cursor on the target and speech for clicking. Consistent with the findings of Halverson et al. [8], the participants who developed this multi-modal strategy tended to have more experience using computers. The participants with limited computer experience may benefit from specific instructions or designs that explicitly encourage the adoption of multi-modal strategies when appropriate.

The fact that participants used multiple applications and frequently switched between applications highlights the importance of keeping the speech dialogue design consistent across applications. This is challenging considering the vast array of commands and tasks that an individual may encounter, but the benefits could be significant.

The first field study indicated that some participants failed to retain the skills they had initially learned with speech-based text entry, and that the adoption of speech technology did not directly correspond to productivity.

Compared to previously reported studies of speech applications, these field studies have the strength of being conducted in real environments. However, this introduced a number of challenges and constraints that made the environment less than ideal. The computers with the speech-based solutions were provided by the researchers. Some participants completed tasks with their own computers, and found it hard to switch to other computers. In some cases, switching computers was not practical due to the use of applications that were not available on the computers we provided. For example, in the first study, eight of the ten participants (all five without disabilities) used both the computer we provided and their own computer. This limited the amount of information that could be collected and also had a potential impact on the interaction styles.

6 Conclusions

The studies provided first hand data on how users with and without physical disabilities used speech applications with personal computers in real environments over a prolonged period of time. The results suggest that participants without disabilities used a greater variety of applications than participants with disabilities, but those with disabilities were more satisfied with the speech-based solutions they experienced. The results also suggest that user requirements for speech-based interactions have evolved

as technology has matured to include better support for formatting text documents as well as more effective support for other applications. The use of speech recognition to support informal communication is particularly interesting. Further research should address these newly emerging requirements for speech technologies.

Acknowledgments. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. CNS-0619379 and National Institute on Disabilities and Rehabilitation Research (NIDRR) under grant number H133G050354. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF and the NIDRR.

References

1. Arnold, S., Mark, L., Goldthwaite, J.: Programming by Voice, VocalProgramming. In: Proceedings of ASSETS 2000, pp. 149–155 (2000)
2. Baber, C., Hone, K.: Modeling Error Recovery and Repair in Automatic Speech Recognition. International Journal of Man-Machine Studies 39, 495–515 (1993)
3. Christian, K., Kules, B., Shneiderman, B., Youssef, A.: A Comparison of Voice Controlled and Mouse Controlled Web Browsing. In: Proceedings of Assets 2000, pp. 72–79 (2000)
4. Dai, L., Goldman, R., Sears, A., Lozier, J.: Speech-Based Cursor Control Using Grids: Modeling Performance and Comparisons with Other Solutions. Behaviour and Information Technology 24(3), 219–230 (2005)
5. de Mauro, C., Gori, M., Maggini, M., Martinelli, E.: Easy Access to Graphical Interfaces by Voice Mouse (2001), Available from the author at: demauro@dii.unisi.it
6. Feng, J., Karat, C.-M., Sears, A.: How Productivity Improves in Hands-Free Continuous Dictation Tasks: Lessons Learned from a Longitudinal Study. Interacting with Computers 17(3), 265–289 (2005)
7. Feng, J., Sears, A.: Using Confidence Scores to Improve Hands-Free Speech-Based Navigation in Continuous Dictation Systems. ACM Transactions on Computer-Human Interaction 11(4), 329–356 (2004)
8. Halverson, C., Horn, D., Karat, C.-M., Karat, J.: The Beauty of Errors: Patterns of Error Correction in Desktop Speech Systems. In: Proceedings of INTERACT 1999, pp. 133–140. IOS Press, Amsterdam (1999)
9. Harada, S., Landay, J., Malkin, J., Li, X., Bilmes, J.: The Vocal Joystick: Evaluation of Voice-Based Cursor Control Techniques. In: Proceedings of ASSETS 2006, Portland, Oregon, pp. 197–204 (2006)
10. Harada, S., Wobbrock, J.O., Landay, J.A.: VoiceDraw:A Hands-Free Voice-Driven Drawing Application for People with Motor Impairments. In: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 27–34. ACM, New York (2007)
11. Hubbell, T., Langan, D., Hain, T.: A Voice Activated Syntax-Directed Editor for Manually Disabled Programmers. In: Proceedings of ASSETS 2006, pp. 205–212 (2006)
12. Kamel, H., Landay, J.: Sketching Images Eyes-Free: A Grid-Based Dynamic Drawing Tool for the Blind. In: Proceedings of ASSETS 2002, pp. 33–40 (2002)
13. Karat, C.-M., Vergo, J., Nahamoo, D.: Conversational Interface Technologies. In: Jacko, J., Sears, A. (eds.) The Human-Computer Interaction Handbook, pp. 169–186. Lawrence Erlbaum and Associates, Mahwah (2003)

14. Karimullah, A., Sears, A.: Speech-Based Cursor Control. In: Proceedings of ASSETS 2002, pp. 178–185 (2002)
15. Manaris, B., Harkreader, A.: SUITEKeys: A Speech Understanding Interface for the Motor-Control Challenged. In: Proceedings of the 3rd International ACM SIGCAPH Conference on Assistive Technologies (ASSETS 1998), pp. 108–115 (1998)
16. Manasse, B., Hux, K., Rankin-Erickson, J.: Speech Recognition Training for Enhancing Written Language Generation by a Traumatic Brain Injury Survivor. Brain Injury 14(11), 1015–1034 (2000)
17. McNair, A., Waibel, A.: Improving Recognizer Acceptance through Robust, Natural Speech Repair. In: Proceedings of the International Conference on Spoken Language Processing, pp. 1299–1302 (1994)
18. Mihara, Y., Shibayama, E., Takahashi, S.: The Migratory Cursor: Accurate Speech-Based Cursor Movement by Moving Multiple Ghost Cursors Using Non-Verbal Vocalizations. In: Proceedings of ASSETS 2005, pp. 76–83 (2005)
19. Pieper, M., Kobsa, A.: Talking to the Ceiling: An Interface for Bed-Ridden Manually Impaired Users. In: CHI 1999, Extended Abstract, pp. 9–10 (1999)
20. Sears, A., Karat, C.-M., Oseitutu, K., Karimullah, A., Feng, J.: Productivity, Satisfaction, and Interaction Strategies of Individual with Spinal Cord Injuries and Traditional Users Interacting with Speech Recognition Software. Universal Access in the Information Society 1, 4–15 (2001)
21. Sears, A., Feng, J., Oseitutu, K., Karat, C.-M.: Speech-Based Navigation During Dictation: Difficulties, Consequences, and Solutions. Human Computer Interaction 18(3), 229–257 (2003)
22. Sporka, A., Kurniawan, S., Mahmud, M., Slavik, P.: Non-speech Input and Speech Recognition for Real-Time Control of Computer Games. In: Proceedings of ASSETS 2006, pp. 213–220 (2006)
23. Zhu, S., Ma, Y., Feng, J., Sears, A.: Speech-Based Navigation: Improving Grid-Based Solutions. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 50–62. Springer, Heidelberg (2009)