

Making Assistive Technology and Rehabilitation Engineering a Sure Bet

Differences in Speech Recognition Use on the Internet

Michael William Boyce

Center for Assistive Technology and Environmental Access - Georgia
Institute of Technology
Atlanta, GA 30318

ABSTRACT

This paper examines the unique aspects of speech recognition software use on the internet. The goal is to answer the question: What can be done to more effectively accommodate speech recognition use on the internet? Six users with disabilities comment and provide task-based explanations on everyday internet activities. Results show a preference for improved mass navigation, integration with web applications and communication interfaces and task focused training methods.

KEYWORDS:

Speech Recognition, Internet Access, Computer Access

BACKGROUND

Speech recognition is one of the most commonly suggested assistive technology solutions for persons with disabilities (1). While there is a plethora of research regarding the effectiveness of speech recognition (2, 3), Mitchard and Winkles note that the use of speech recognition varies according to the desired task (4). If a user cannot complete their desired tasks, there is a greater chance that they will abandon the product. In fact, with specific regard to speech recognition, there still remains a high tendency of abandonment (5, 6).

One reason for abandonment, and a focus of most speech recognition research, is the problem of words and commands not being correctly recognized. IBM, in partnership with Google, is currently working on testing computer systems that can perform better recognition with regard to human speech input as a part of their Super Human Speech research (7).

Much less research exists on the use of speech recognition with integrated web applications or when navigating the internet. Current internet user commands built into speech recognition technology are primarily for conducting searches. WebSpeak, is one tool that is used to help speech recognition users access other commands. The customizable speech-based web navigation interface includes a mode that restricts the user to the list of available words on the page, improving recognition for those words (8).

However, with the increased use of multipurpose design, the applications in which speech recognition may be used is consistently growing more complex. Speech recognition users now engage in social networking, web based services, and entertainment, to name a few. This research study strove to discover the complexities and challenges related to the use of voice input with these types of applications.

Making Assistive Technology and Rehabilitation Engineering a Sure Bet

RESEARCH QUESTION

The focus of this study is to determine how to more effectively accommodate speech recognition use on the internet.

METHODOLOGY

Procedure: In an effort to allow flexibility to research subjects and consistency between responses, interviews were chosen as the method for data collection. Six subjects took part in the interviews which were semi-structured and facilitated by phone. The subjects were recruited via the CATEA Consumer Network (CCN) as well as through Peer Support at the Shepherd Spinal Center. The subjects gave oral consent prior to participating in the study and were not compensated.

The interview included questions about their speech recognition usage, functional limitations, and the type of internet sites they use. Subjects were also asked about their experience of browsing using speech recognition and how they would go about completing various internet tasks using speech recognition. The tasks were: Ordering something from an online store, catching up on the latest news stories, filling out a form for customer support, checking email, chatting with friends via instant message, social networking sites, and virtual environments (such as Secondlife).

These tasks were chosen through an informal brainstorming session with individuals with disabilities to understand what tasks they perform online. They were given the option to skip any activities that they didn't have experience with.

RESULTS

Four of the subjects were male and two were female; all were over the age of 18 at time of the interview. Quadriplegia with limited use of one hand was the most prevalent (3 out of 6) functional limitation. There was also one patient with tetraplegia with the use of one hand, one with dyslexia, and one with rheumatoid arthritis.

All of the subjects used a version of Dragon speech recognition software, and four of the six use speech recognition for all computer related activities. The most predominant software was Naturally Speaking version 9 (3 of 6), followed by not sure (2 of 6) and Dragon Dictate 3.0 classic. All subjects used voice input along with a combination of keyboard and mouse. Three of six participants use a combination of voice, mouse, and keyboard. Two of six participants use voice and mouse, while the remaining participant uses voice and a mouth stick to control keypad mouse commands.

All of the subjects used speech recognition for at least a portion of their web browsing, Four of the six participants explained that they access the internet using voice because they don't have the hand function to accurately control a mouse, complete emails quickly, and word correction. The other two participants used speech recognition for data entry on internet sites, but found it more efficient to navigate the internet using other means. With an average user satisfaction rating of 3.6 on a scale of 1 to 5, reasoning for rating came from two sources for 4 of the 6 users: time consumption and command recognition. Time consumption was specifically linked to the process of navigation and effectively moving around and between sites. One participant even went so far as to not update from an older

Making Assistive Technology and Rehabilitation Engineering a Sure Bet

Application	Is Command Recognition Supported to Access Menus?	Does The System recognize words in Composition?	Does the system recognize words in Chat Interfaces?	Does the System Support Mass Navigation?
Outlook	Yes	Yes	N/A	Yes
Word	Yes	Yes	N/A	Yes
Web Blogs	Yes	Marginal	N/A	N/A
Gmail	Yes	Yes	Marginal	Yes
Facebook	Yes	N/A	Marginal	Marginal
Amazon	Marginal	N/A	N/A	Marginal
IM	Yes	N/A	Yes	N/A

Table 1: Speech Recognition Capabilities As Reported From Participants

version of Dragon, in order to retain the MouseGrid command which allows for the screen to be broken up into nine boxes for selection. Also, two subjects mentioned increased time consumption as their voice changes throughout the course of the day due to the fact that tonality changes and strain from fatigue can cause discrepancies between the saved voice files and the users voice, impeding recognition.

One of the problems noted was the inability for speech recognition to understand the difference between a command and conversation chat in Gmail and Facebook. As such, conversation could be interrupted by a misinterpreted command that closed the window. Also, with regard to gaming, all users found it very challenging, if not impossible, to participate in both stand alone and add-on type games due to the lack of command support for voice input. For more information on interaction experiences with different applications see table 1.

Each of the participants expressed a lack of proper guidance and minimal training to begin using the program. This often led to either not understanding important interface options (4 of 6) or the need to search for more advanced features independently (2 of 6). The two individuals who did investigate further learned how to use macros, which helped with internet tasks such as form data entry.

DISCUSSION

The relatively small sample size means that the data, while indicative, are not conclusive but can help in informing the direction of future developments. For future developments of Dragon from a design perspective, changes can go a long way toward assisting the user. Improving the navigation and focus aspects of voice input could help the user navigate the interface. For example having a way to activate a free flowing mouse where the mouse continuously moves as directed by voice rather than discrete commands. Although MouseGrid functionality exists in Dragon 10, users should be taught this and other features that they may not discover independently. From a usability perspective, it would be beneficial to incorporate “non-professional” based command libraries that can be installed to assist with areas such as social networking, gaming, and chatting to accommodate to the greater needs of persons with disabilities.

Making Assistive Technology and Rehabilitation Engineering a Sure Bet

There are several things that designers of internet-based applications could also do to make their sites more compatible with speech recognition tools. Site commands and keywords can be chosen to be consistent with those already built into speech recognition commands. In addition, building in the ability to activate the commands with keyboard controls will enable speech recognition users to activate those commands through macros. Finally it is important to examine how technology can be built into web design to facilitate easier interaction with voice input. It could be a matter of attaching a voice framework certain parts of an HTML document to recognize voice commands.

For service providers, it appears that the consumers are placed into a situation where the product is purchased and the users either receive very little training or are left to fend for themselves entirely. As noted in (9), this requires an intense motivation and need for the individual user in order for successful adaptation to disability and the activities of daily life. One of the possibilities to consider is using mouse / keyboard control methods to serve as a backup to voice. Furthermore it appears that many users are unfamiliar with the use of macros. If person to person training is necessary, determine what tasks would be most difficult for the user to learn independently, such as macros, and focus training around those activities.

As the footprint of technology continues to change to accommodate for everyday interactions, so must speech recognition technology. This includes applicability to mobile devices for which new research is currently underway (10). By focusing on educating the users on the appropriate techniques and options available to them, persons with disabilities will be able to more efficiently use speech recognition systems as the internet continues to develop.

ACKNOWLEDGEMENTS

This study was conducted as part of the RERC on Workplace Accommodations, which is supported by Grant H133E070026 of the National Institute on Disability and Rehabilitation Research of the U.S. Department of Education. The opinions contained in this publication are those of the grantee and do not necessarily reflect those of the U.S. Department of Education.

REFERENCES

1. Gamble, M.J., D.L. Dowler, & Orslene, L. (2006). Assistive Technology: Choosing the right tool for the right job. *Journal of Vocational Rehabilitation*, 24(2): 73-80.
2. Jutai, J., Coulson, S., Fuhrer, M., Demers, L., & DeRuttyer, F. (2008). Article 5:predicting assistive technology device continuance and abandonment . *Archives of Physical Medicine and Rehabilitation*, 89(10): e2. doi: 10.1016/j.apmr.2008.08.013
3. Kalinli, O., Seltzer, M.L., & Acero, A. (2009). Noise adaptive training using a vector taylor series approach for noise robust automatic speech recognition. *Proceedings of the 2009 IEEE international Conference on Acoustics, Speech and Signal Processing*, 3825-3828. doi: <http://dx.doi.org/10.1109/ICASSP.2009.4960461>

Making Assistive Technology and Rehabilitation Engineering a Sure Bet

4. Mitchard, H. & Winkles, J. (2002). Experimental Comparisons of Data Entry by Automated Speech Recognition, Keyboard, and Mouse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44(2): 198-209.
5. Koester, H.H. (2003). Abandonment of Speech Recognition Systems by New Users. Proceedings of RESNA 2003 Annual Conference, Atlanta, GA. Arlington, VA: RESNA Press.
6. Verza, R., Lopes Carvalho, M.L., Battaglia, M.A., & Messmer Uccelli, M. (2006). An Interdisciplinary approach to evaluating the need for assistive technology reduces equipment abandonment. *Multiple Sclerosis*, 12(1): 88-93.
7. Howard-Spink, S. (2002, September 18). You just don't understand!. IBM Think Research, Retrieved December 10th, 2009 from http://domino.watson.ibm.com/comm/wwwr_thinkresearch.nsf/pages/20020918_speech.html
8. Hamidi, F., Spalteholz, L., & Livingston, N. Web-speak: a customizable speech-based web navigation interface for people with disabilities. Retrieved January 14, 2010 from <http://www.cse.yorku.ca/~fhamidi/resources/WebSpeak.pdf>
9. Roberts, K.D., & Stodden, R.A. (2005). The use of voice recognition software as a compensatory strategy for postsecondary education students receiving services under the category of learning disabled. *Journal of Vocational Rehabilitation*, 22. Retrieved December 10th, 2009 from https://scholarspace.manoa.hawaii.edu/bitstream/10125/9024/1/uhm_phd_4362_r.pdf
10. AT&T/Vlingo bring speech recognition to mobile devices. (2009, October 1). Audiotex Update, Retrieved December 10th, 2009 from http://www.thefreelibrary.com/*AT%26T%2fVLINGO+BRING+SPEECH+RECOGNITION+TO+MOBILE+DEVICES.-a0208031119