https://doi.org/10.48009/2_iis_2007_541-546

The Potential Impact of Speech Recognition Technology On Workplace Productivity

Hal Records, PhD, Bryant University; Nancy M. Records, MS, MBA, Bryant University; Richard Glass, PhD, Bryant University; Robert Behling, PhD, Arrowrock Technology; Janet Prichard, PhD, Byrant University

ABSTRACT

The effect of desktop applications programs on knowledge worker productivity has been significant and well documented. However, for the most part technology driven productivity gains have been at a standstill for last few years. Recent modifications to office suites and operating systems have been more superficial than substantive, and with the exception of improved search and data security capabilities, software introductions have not substantially advanced worker productivity. That may be about to change. Automatic speech recognition (ASR) software embedded in popular word processing programs shows great promise, offering the potential for a faster and more effective human computer interface.

A study was conducted to investigate 1) Is voice recognition ready for prime time; and 2) What is the learning curve for desktop voice recognition technologies. It was found that the user-friendliness of ASR was perceived as being very good, that the likelihood of future use of ASR was very high, and that user comfort levels with dictation, editing and accuracy were tepid. This suggests that lack of familiarity may contribute to user hesitation in adoption of ASR technology, but there is a recognition and willingness among users to pursue it.

Keywords: Workplace productivity, personal productivity software, speech recognition, voice recognition, human computer interface, keyboard alternatives, workplace design

INTRODUCTION

Automatic speech recognition (ASR) is technology that allows a computer to map a stream of acoustic signals into a sequence of words. Applied to the desktop ASR allows users to enter text and commands by speaking into a microphone. Although speech recognition is not new, recent advances in high speed multi-core desk top processors, storage capacity and lower costs have made the technology more reliable and accessible to the public. Software algorithms have improved dramatically, increasing speech recognition accuracy by as much as 50% in just the last year.

In spite of the improvement in technology, the market has been slow to embrace ASR. Gartner (2005) claims that there is an image problem with ASR. ASR has not experienced the 'hot and sexy period" that other new technologies have benefited from. Bad press regarding ASR training and accuracy problems abound (Financial Times, 2006). Most recently, a senior Microsoft executive introduced the new ASR capabilities of Vista in a live demonstration. The comical but sad failure of the system during the demonstration was captured and broadcast to millions of people over YouTube.

There is evidence that market perceptions regarding speech recognition products, functionality, and ease of use are beginning to change. In 2006 the overall market for ASR technology topped 1 Billion dollars, an increase of 100% in just two years [2] and sales are projected to double again by 2009. The public is gaining greater exposure to ASR through the major commitment by Microsoft to provide the technology free with the Windows operating systems and Microsoft Office. Initial market place reviews of ASR in Vista suggest that it is a large improvement over Windows XP and that it functions very well [12]. Microsoft's entry into the field has sparked interest in software developers to embed ASR in a host of applications and has raised the bar for the competition. Dragon Naturally Speaking has just introduced version 9 of their software with a large array of improvements. We may well be seeing the beginnings of the major breakthrough in market dynamics that Gartner claims will be necessary for ASR to become a mainstream technology.

Effective use of speech recognition technologies on the desktop may well be the impetus for true productivity gains for knowledge workers in the future. However, speech recognition on the desktop is likely to be accompanied by changes in user expectations, work procedures, behavior and outputs. Users will need to invest time "training" their speech recognition software, will need to build speed and confidence, and will need office space and furniture designed to control noise and filter sound.

This paper reports the results of the first of two studies designed to gain a better understanding of the current state of ASR for word processing activities. The objective of the first study is to explore first time user initial perceptions of the ease of set up and use of ASR and the likelihood that the first time user would use ASR in the future. The second study will address the performance of ASR compared to more traditional keyboard and mouse data entry. The first study addresses two questions: 1) Is speech recognition ready for prime time; and 2) What is the learning curve for desktop speech recognition technologies? This understanding will aid in developing and delivering training for utilization of these technologies, as well as providing a better understanding of the challenges faced by both organizations and individuals when these technologies are deployed. The contribution of this study lies in gaining a better understanding of the acceptance and implementation of desktop voice recognition technologies.

LITERATURE REVIEW

Speech recognition engines have been available for several decades. AT&T Bell Laboratories developed a system in 1952 that could recognize digits between 0 and 9 spoken over a telephone. Their claim of 98% accuracy inspired Hyde's Law which facetiously stated (reported in Vinciguerra and Kun, [12] that "Because speech recognizers have an accuracy of 98%, tests must be arranged to prove it." Today, claims of accuracy for ASR still hover around the same 98% for complex language sequences and much like the tests of 1952; the results are still dependent on the test design. This results from the fact that several factors influence the accuracy of ASR including: (1) the size of the vocabulary that ASR needs to recognize, (2) how fluent, natural or conversational the speech is, (3) variation in channel and noise and (4) speaker-class characteristics that include dialect, foreign accents, age and similarity to the data the system was trained on (Martin, 2005).

Furui [5] identifies several sources of acoustic variation in speech that stem from: (1) speaker's voice including quality, pitch, gender, dialect, (2) Speaking style including stress/emotion, speaking rate, Lombard effect, (3) Task/Concept including man-machine dialogue, dictation, free conversation, interview and (4) phonetic/prosodic context. Zohar [12] claims that more than 50% of people who try ASR do not continue to use it because of despair with the system. He recommends that these individuals need to learn about their own use of language before trying to use ASR.

Nonetheless, in a properly controlled environment that makes use of a good microphone, quiet location, dedicated dictionary, skilled language practitioner and a well trained engine, high levels of accuracy are attainable for ASR word processing applications [12]. If high accuracy rates are possible then ASR offers several advantages over keyboard data entry. The main advantages include: hands free entry of data, reduced upper limb fatigue, increased speed of data entry particularly for those with poor typing skills, use of macros to automate long and/or complex data entry sequences or to provide guided feedback, greater security and lower costs and higher productivity.

To date users have been resistant to adopting ASR technology [1,7,8,9, Costanza, 2003]. Even though ASR technology is bundled free of charge with Microsoft Windows and may be used in all Microsoft Office applications, most users have never attempted to use the ASR technology. In order for ASR to achieve its potential for increasing the productivity of desktop applications more users will have to try the application and adopt it for use on a regular basis. The technology acceptance model (TAM) developed by Davis [4] has become widely used in the literature to predict whether a user will or will not adopt a new technology. The TAM model suggests that a user's perceptions of the usefulness and ease of use of the new technology influences attitudes towards the new technology which in turn influences the user's behavioral intention to adopt the new technology. Davis suggests that behavioral intention to use a technology is the prerequisite to actual system use.

This exploratory study is designed to shed light on the attitudes of first time users of ASR and to investigate whether the users are receptive to using the technology again. More specifically, the users' perceptions of the ease of set up and use of the system are measured. User perceptions of the system's performance are also measured and the users' attitudes towards the system and likelihood of using the system again are explored.

METHODOLOGY

The two questions: 1) Is speech recognition ready for prime time; and 2) What is the learning curve for desktop speech recognition technologies are addressed through the use of a survey questionnaire and a series of exercises administered to undergraduate students enrolled in an introductory level technology course at a Northeast University. While the survey population is not representative of a cross section of the American workforce, these students will shortly become the new business professionals, have grown up with rapid technological change, and are among those most likely to adopt a new technology.

All students participating in the exercise and survey used their own high-end IBM Thinkpad laptops and ASR software contained in Word 2003 and Windows XP (updated with recent service packs). Hence hardware and software platforms were identical for the120 students who completed the exercise. Students were asked to activate speech recognition, train the computer to recognize their voices, dictate ten brief phrases contained in the exercise including their first and last names, and complete a 10 question survey based on their experience with the exercise. Most students utilized microphones built into their laptops, but a few used headset microphones.

Survey questions are shown in Figure 1. A number of students wrote comments in the space provided on the bottom of the questionnaire.

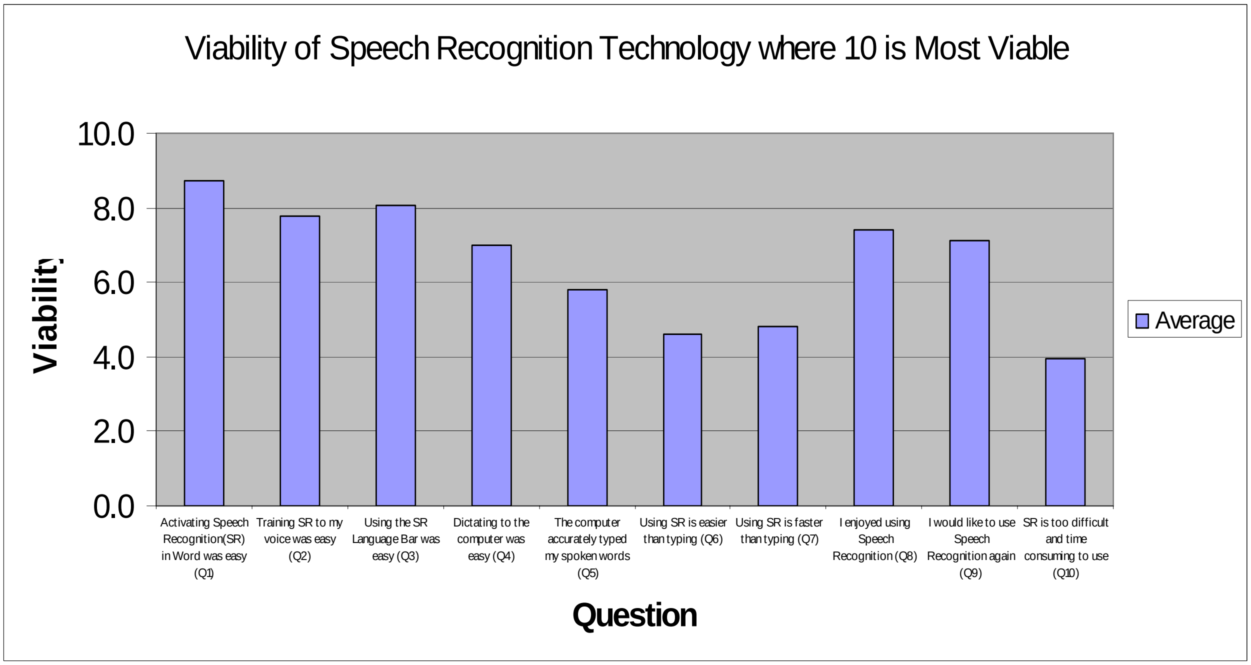

Figure 1. Viability of Automatic Speech Recognition Software

Questions 1,2 and 3 were designed to measure user friendliness of the ASR software including ease of activating (1), training it to recognize the users voice (2), and using the speech recognition Language Bar (3). Questions 4,5,6 and 7 were designed to measure user comfort levels with dictating to the computer (4), getting accurate results from dictation (5), using ASR as opposed to the much more familiar keyboard (6), and increasing speed versus keyboard entry(7). Questions 8,9 and 10 were designed to measure the likelihood that respondents would continue to use ASR technology in the future. Question 8 indicates to what degree users enjoyed using the ASR, question 9 the likelihood that they will use it in the future, and question 10 an inverse validation designed to trap arbitrary questionnaire completion. It is interesting to note that survey results appear to contain little or none of this.

FINDINGS

Figure 1 shows the average value for all 120 respondents for each question on a scale of 1 to 10 where 10 is most viable. User-friendliness of ASR was perceived as being very good with over-all values of 8.7 for ease of activating the program, 7.9 for training, and 8.0 for using the language bar. This aggregates to an average 8.2 for those three questions representing user friendliness.

Figure 2 shows the number of respondents and their evaluation of ease of activating the program, ease of training, and ease of using the language bar. As shown by Figure 2, activating ASR software was considered the easiest task, followed by using the Language Bar and then training. It is interesting to note how few considered these tasks challenging and how many considered them to be very easy.

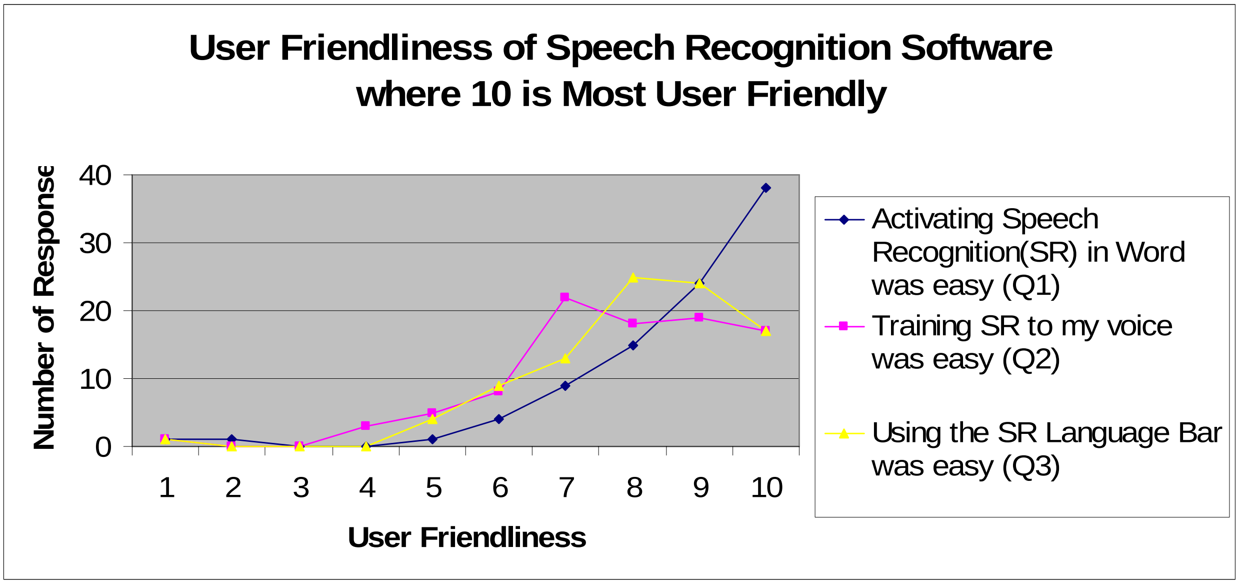

Figure 2. User Friendliness of Automatic Speech Recognition Software

Figure 3 shows the number of respondents and their comfort level with the new ASR technology. Response to question 4 indicates that dictating to the computer is easy. Response to question 5 shows a mixed mid range response to the accuracy of the program. A number of respondents indicated that they believed more training time would improve accuracy as would more practice with dictation. Response to question 6 received a largely unfavorable response. Users evidently believe that the keyboard is easier to use than dictation, which may in part be do to the fact that this was the very first time many had tried ASR. Likewise users believe that dictation is slower than typing and many commented that the correction of errors was considerably more time consuming than the original dictation. Again, this may be an indicator of unfamiliarity with the program, and may suggest that keyboard error correction after dictation may be a useful practice, and way to improve speed.

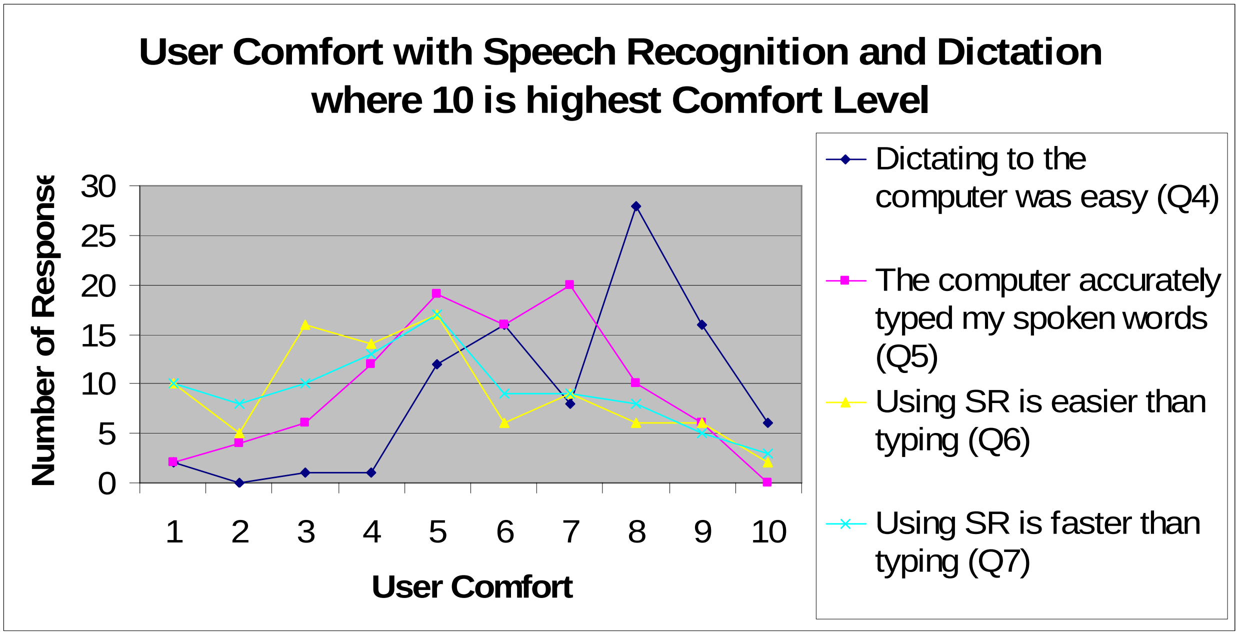

Figure 3. User Comfort with Automatic Speech Recognition Software

Figure 4 shows the likelihood of future use of ASR. Response to questions 8 and 9 indicate a very strong likelihood of future use. When viewed in light of the TAM model this outcome suggests that users' attitude towards this new technology is the precursor to actual system use.

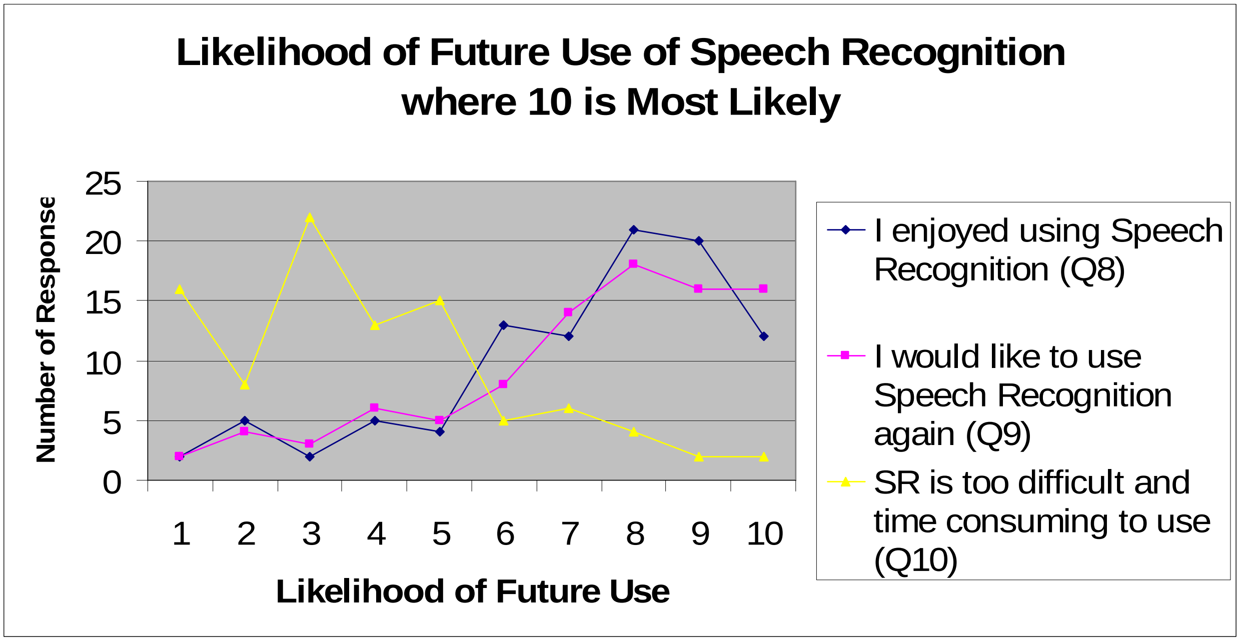

Figure 4. Likelihood of Future Use of Automated Speech Recognition Software

CONCLUSIONS AND IMPLICATIONS

User-friendliness of automated speech recognition technology is perceived as being very good, the likelihood of future use by those doing word processing is very high, and user comfort levels with dictation, editing and accuracy were tepid. These findings indicate that speech recognition may indeed, after many decades, be ready for prime time. Study findings also indicate that the learning curve for speech recognition is relatively short and not terribly steep, but that users do not have proficiency in its use or confidence in its accuracy or speed. Comments by users suggest that additional training of the software, more practice with dictation, keyboard use for post dictation error correction, the advent of high powered dual core processors, and attention to the creation of quiet dictating environments may serve to mitigate the remaining obstacles to wholesale adoption of this technology and to a subsequent surge in knowledge worker productivity.

Subsequent to completion of this paper the same survey was administered to a group of approximately 40 graduate students whose average age is nearly 10 years above the original survey group. It is interesting to note anecdotally that their general response was less favorable, which suggests that upcoming generations may be more predisposed to using ASR.

Utilization of speech recognition on the desktop is likely to force changes in user expectations, work procedures, behavior, outputs and workplace design. Given the likelihood of future use of ASR as indicated by this study, and the significant potential for increased knowledge worker productivity, a second study is planned that will specifically address what techniques, training and physical office space facilities will be required to realize the productivity potential of automatic speech recognition programs.

REFERENCES

1. Anson D. et. al., (2004). Does Speech Recognition Deserve Recognition? A Study Comparing the Efficacy of Speech Recognition vs. the Standard Keyboard. *Assistive Technology Research Center*, Dallas PA, Misericordia College.
2. Borzo, J. (2007). Now You’re Talking. Business 2.0 Magazine, February, 2007.
3. Davis, C. (2001). Automatic Speech Recognition and Access:20 Years, 20 Months or Tomorrow. *Hearing Loss*, 22(4) 11-44.
4. Davis, F. D. (1986). A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. Unpublished doctoral dissertation, Sloan School of Management, MIT.
5. Furui, S. (2000). Automatic Speech Recognition and its Application to Information Extraction. http://www.furui.cs.titech.ac.jp/
6. Grasso, M. (1995). Automated Speech Recognition in Medical Applications. *Automated Speech recognition in Medical Applications*. 1-8.
7. Green, H. D. (2003). Speech Recognition Technology for the Medical Field. *Journal of American Academy of Business, Cambridge* 2,2. 299-303.
8. Koester, H. Horstman (2004). Usage, Performance and Satisfaction Outcomes for Experienced Users of Automatic Speech Recognition. *Journal of Rehabilitation Research and Development* 41, 5. 739-754.
9. Simon, A. J., Paper, D. (2007). User Acceptance of Voice Recognition Technology: An Empirical Extension of the Technology Acceptance Model. *Journal of Organizations and End User Computing* 19(1), 24-50.
10. Thomas, K. (2006). Voice Recognition Software Starts to Make a Big Noise. *Financial Times*, June 14, 2006.
11. Vinviguerra, B. J., Kun, A. L. (2003). A Comparison of Commercial Speech Recognition Components for Use in Police Cruisers. Electrical and Computer Engineering Department, University of New Hampshire.
12. Zohar, I. (2007). Speech Recognition. http://www.tau.ac.il/~itamarez/sr/index.html