Tech Talk: "You Talk, it Types?" Not Quite: Speech Recognition Technology for Postsecondary Students with Disabilities

David McNaughton
The Pennsylvania State University

BACKGROUND

For all of us who have struggled with a computer keyboard, the allure of creating text simply by speaking is strong. For individuals with disabilities, alternatives to the regular keyboard *are especially important. Individuals with physical disabilities may experience difficulty in performing the rapid discrete movements necessary for efficient typing. Individuals with learning disabilities may experience difficulty both in learning keyboarding skills, and in managing the cognitive load associated with simultaneously generating ideas, spelling words correctly, and entering text accurately.*

Since the 1960s, computer manufacturers and software designers have worked to develop speech recognition technology that would provide computer access for individuals who cannot type efficiently. More recently, the computer industry has targeted the development of efficient and easy to use speech recognition systems as a top priority. In a 1997 speech to software developers, Microsoft President Bill Gates (September, 1997) commented that "It's my belief that every computer will have speech and linguistics built into it, because that's the world of discourse that we live in." The computer industry hopes that removing the keyboard "barrier" will provide greater access to computer technologies.

The development of a new generation of speech recognition technology, "continuous" speech recognition, has promised to bring the dream of effective and efficient speech recognition systems a step closer to reality. Traditional "discrete" technology required the individual to speak in individual words separated by pauses. The new continuous technology promises to recognize "connected" (i.e., conversational) speech. This column discusses the strengths and weaknesses of current speech recognition technology, as well as the implications of its use with postsecondary students with disabilities.

Speech Recognition Technology

"Speech recognition technology" is designed to do just that-enable an individual to control a computer using only speech. Rather than using a keyboard, the user can tell the computer (using a microphone) to execute commands ("Open file `Leaving"') and enter text (e.g., "Dear John").

Discrete Speech Recognition

Until recently, all major speech recognition systems used "discrete" technology. For these systems, individuals need to learn to speak in a clearly articulated word-by-word manner, with brief pauses between words, so that their speech will be recognized by the computer. In addition, individuals need to "train" the computer to recognize their particular pronunciation of words. As the user makes use of the system, the speech recognition software develops a user-specific voice-file that contains information about the user's voice qualities and pronunciations. This information is used by the computer to make guesses about the words used by the individual.

Because speech recognition can never guarantee 100% accuracy, the computer generates a list of alternative words each time it makes its "best guess" as to what the user has said. If the computer was inaccurate with its original guess, the user can select the desired word from the list of alternatives. The user can also choose to spell out the word letter by letter using speech, or by using the keyboard. A very good demonstration of speech recognition technology is available at the National Center for Improvement in Practice (NCIP) Voice Recognition demonstration page.[1]

Leading speech recognition systems using discrete technology are Dragon Dictate[2] and IBM VoiceType[3] for Windows users, and Power Secretary for Macintosh users. The average rate for typical adults after training is between 45 and 65 words per minute.

Investigations of the use of this technology by postsecondary students with learning disabilities has produced mixed results. While Higgins and Raskind (1995) documented evidence of a powerful positive impact for the college students with whom they worked, Day (1995) reported that speech recognition had little overall impact with her students. It is unclear how variables related to student characteristics, software use, targeted writing activities, or training activities may have contributed to the differences in these findings.

Continuous Speech Recognition

During the past year, the introduction of continuous speech technology has dramatically changed the field of speech recognition. This approach allows the user to dictate by speaking in a normal, continuous, and conversational manner. Continuous speech systems are also significantly cheaper than the older discrete systems, with state of the art technology available for approximately $150.

These new systems also promise to dramatically reduce the training time necessary for initial use. Because the individual can use normal everyday speech, the user does not learn to speak using individually distinct words. The manufacturers have also promised that these new technologies provide high rates of recognition "out of the box," so that the time necessary for the computer to learn to recognize your voice is reduced.

Leading speech recognition systems using discrete technology are Dragon Naturally Speaking[5] and IBM ViaVoice[6] . A detailed performance comparison of these systems is available at the 21st Century Eloquence Information in the News site; it was the opinion of reviewers that NaturallySpeaking enjoyed a slight advantage over ViaVoice for both ease of use and recognition accuracy. With training and extensive use, these systems recognize the user's speech (at about 80 words per minute) with approximately 90-95 % accuracy.

Although continuous speech recognition software presents many new promising features for individuals with disabilities, it also provides a number of serious challenges (NCIP Spotlight on Voice Recognition). First, recognition is enhanced when the user speaks in complete sentences; this method of text composition may be challenging for some users. Second, the user may be at the end of their dictated sentence before the words from the beginning of the sentence appear on the screen, which can pose problems for individuals who are monitoring their performance. Third, error correction may necessitate changes to whole passages of text, not just individual words.

Factors Supporting Effective Use

Disability support service providers have noted that the successful use of this technology is influenced by two key factors: the attitude of the user, and the successful use of training and practice activities.

Attitude of the user.

Speech recognition software is used most successfully when an individual is prepared to make a serious commitment to the training necessary to achieve proficiency. The likelihood of a positive reaction and commitment to the thorough investigation of the software is highly dependent on three factors.

First, the trainer should ensure that the system is set up to function at a high level of accuracy. "Slow" computers, poor sound cards, and improperly positioned microphones will greatly degrade performance.

Second, the trainer should present the software in an incremental manner that is paced to the user's developing understanding. The overall computer literacy of the individual is an important consideration at this time. As noted by Randy Borst (post to DSSHE-Listserv, June 6, 1996) "that's part of the rub with access computer technology-most all of it requires a longer learning curve and a better command of the access technology than is needed for the application program itself."

Third, the trainer must consider the student's personal desire to make use of the software. Students who do not feel that they need a new approach to keyboarding, and who feel that they have been "coerced" (by family members or other professionals) into trying the new software, may never "buy into" giving the program a chance. This is usually very evident from the effort students demonstrate during training activities, and the amount of time they spend between training session practicing what they have been taught (J. Kennedy, personal communication, May 22, 1998).

Training activities. Despite the claims for outstanding "out of the box" performance, anyone undertaking the use of speech recognition software should understand the time commitment necessary to pursue proficiency, While the initial training to master basic operation of continuous speech systems may require less time (approximately 3-6 hours) than the discrete systems (10 - 15 hours) a serious commitment to both training and practice activities is critical for both (J. Kennedy, personal communication, May 22, 1998).

A detailed listing of suggested training activities is available at the NCIP site on voice recognition. In summary, there are two aspects of training with speech recognition. First, the computer must be trained to recognize the user's words. As the system becomes accustomed to the user's voice, it develops a model that enables it to make better predictions as to what the user is saying. Second, all users must be trained in the process of saying words and correcting any mismatches between the user's spoken word and the computer's prediction.

For individuals who will rely upon the speech recognition system for all aspects of computer control, students will also need to learn how to spell by voice, and to give commands to the computer.

Expecting perfection in the early stages of usage is unrealistic, and new users should be encouraged to approach the system with the understanding that 80 - 90 percent recognition at first is part of the expected learning curve. It is important that users continue to use their normal speaking voice without straining, and take occasional drinks of water. Straining your voice in pursuit of "perfect" recognition can cause damage to the vocal cords (Kambeyanda, Singer, & Cronk, 1997). Users must also consistently correct any mismatches produced by the computer. Failure to correct recognition errors results in a "corrupted" voice recognition file; the speech recognition system will be operating on faulty information and recognition may actually decline if errors are left uncorrected.

Summary

Like all technology, voice recognition is not a panacea, and those who believe it to be 'the answer' for everybody (or even anybody) without taking a hard look at its limitations are being just as foolish as those who dismiss it without considering its potential benefits (T. Carpenter, NCIP, 1998). [10]

Speech recognition is a key area of development for mainstream software developers, and individuals with disabilities will undoubtedly benefit from the huge investments currently being made in the development of this software. New speech recognition products must be investigated with care, however, as there may be a large gap between what is presented in advertisements, and the individual's initial experience with the system.

Acknowledgments and Resources

I would like to thank the following individuals who made substantial contributions to this column: Joyce Kennedy (University of Southern Maine) and Ross Pollack (Manhattan College, NY). For those individuals who have questions about the use of speech recognition for individuals with disabilities, the following WWW sites may be of interest. The NCIP Voice Recognition website[11] provides a comprehensive introduction to current technologies, as well as a collection of recent postings to a newsgroup focused on educational applications of speech recognition technology. The Speech Recognition Website[12] provides reviews of voice recognition/dictation software, as well as how-to articles and tips. You can also review discussions on the Speech Recognition Bulletin Board, an on-line, threaded bulletin board for the Speech Recognition community. 21st Century Eloquence provides sales and training support for major speech recognition products (their motto is "Typing is so 20th century") and they provide discounts for students and educators; their website[13] provides links to current industry information and to an online collection of current media reviews of speech recognition software.[14]

References

Day, S. L. (1995). *Computerized voice recognition system effect on writing skills of community college students with learning disabilities.* Unpublished doctoral dissertation, The Florida State University.

Gates, B. (September, 1997). *Remarks to Professional Developers Conference San Diego, CA*.

Higgins, E. L., & Raskind, M. H. (1995). Compensatory effectiveness of speech recognition on the written composition performance of postsecondary students with learning disabilities. *Learning Disability Quarterly, 18*, 159-174.

Kambeyanda, D., Singer, L., & Cronk, S. (1997). Potential problems associated with use of speech recognition systems. *Assistive Technology 2*,95-101.

1 The URL for the NCIP speech recognition demonstration is http://www.edc.org/FSC/NCIP/VR/VR-Demo2.html

2 The URL for Dragon Dictate is http://www.dragonsys.com/frameset/product-frame.html

3 The URL for IBM VoiceType is http://www.software.ibm.com/is/voicetype/us-gold.html

4 The URL for Power Secretary is http://www.dragonsys.com/products/powersecretary.html

5 The URL for Dragon Naturally Speaking is http://www.dragonsys.com/products/natspeaking.html

6 The URL for IBM is ViaVoice is http://www.software.ibm.conVis/voicetype/us_vvg.html

7 The URL for the 21st Century Eloquence Information in the news site is http://cnet.conVContent/Reviews/Compare/Speech/ss03.htmI

8 The URL for the NCIP Spotlight on Speech Recognition site is http://www.edc.org/FSC/NCIP/VR/VR-home.html

9 The URL for the NCIP speech recognition training site is http://www.edc.org/FSC/NCIP/VR/VR-HTMLdoc.html#anchor87777

10 The URL for the NCIP About Spotlight on Voice Recognition site is http://www.edc.org/FSC/NCIP/VR/VR-AboutVR.html

11 The URL for the NCIP Spotlight on Speech Recognition site is http://www.edc.org/FSC/NCIP/VR/VR-home.html

12 The URL for the Speech Recognition Website is http://www.scottrell.com/Speech/index.html

13 The URL for 21st Century Eloquence is http://www.voicerecognition.com/

14 The URL for Links to voice recognition news is http://www.voicerecognition.com/1998/information/in-the-news.html

On The Net

Daniel J. Ryan
State University of New York at Buffalo

The Internet has been said to have democratized information to a great extent. It was the Internet which was the medium by which the world became aware of the massacre in Tiannemen Square in Beijing. It was on the Internet that the Clinton-Lewinsky scandal

became widely known. It was also the Internet which disseminated news of the death of Bob Hope: News which Bob Hope strongly denied. The value of the Internet, and the information it delivers, is only as high as the credibility of those who post the information.

A good example of the value of electronic discussion groups, or listservs like DSSHE-L, the list for Disability Support Services in Higher Education, is the timing and quality of the discussion of the Boston University Learning Disabilities case.

Discussion of the BU case was begun by Ron Blosser of the University of California at Irvine. In February 1996, he called the list's attention to an article that had appeared in that morning's New York Times. Around 15 hours later (speaking to the amount of free time DSS professionals have), Mary Hurwitz (Suffolk University) pointed out that Loring Brinckerhoff and Kip Opperman had both resigned over the situation in question.

Things got going again in September 1996 when a posting to the listserv passed along a letter from Anne Schneider, a mother of a Boston University student, who was instrumental in raising substantial funds for the Boston University program. The letter by Ms. Schneider outlined the current status of the BU lawsuit along with an appeal for financial contributions to help fund the out-of-pocket expenses totaling $85,000.

DSSHE-L users, not immune from the urge to spin-doctor, pointed things out when Boston University was in the headlines twice again in 1997 for different sorts of legal problems. Marcia Carlson (University of Wisconsin) contributed an excerpt from the daily electronic version of "Chronicles of Higher Education" which states that Boston University was ordered by Massachusetts highest court to pay a local training company more than $5.7 million for breach of contract and deceptive trade practices. In October 1997, Randall Borst (SUNY at Bufflao) shared a few details about a settlement where Boston University agreed to pay more than $700,000 to settle charges stemming from a series of environmental mishaps.

Richard, a student from the University of Massachusetts, became a part of the on-line discussion. Being familiar with John Silber (Boston University), and various comments made by him over the years, Richard felt that Mr. Silber was constantly putting his foot in his mouth. He shared an anecdote that exemplified his stance.

When the BU decision was released, all 53 pages of it were quickly posted to the DSSHE-L for anyone interested. Other related information was shared over the listserv including articles which could be found in the Wall Street Journal (September 3, 1997), the September 1997 issue of the Disability Compliance for Higher Education, and the October 9, 1997, issue of USA Today. Jane Jarrow (Disability Access Information and Support) attempted to offer a more positive and proactive response to the Boston University decision by organizing volunteers to create a "press kit" for service providers to use.

The talk of Boston University's legal problems has continued to be a topic for discussion on the DSSHE-L. The final decision handed down by Judge Saris from the United States District Court in Massachusetts, was recently being analyzed by members of the listserv and shared with one another. Kip Opperman (KippOpp Consulting) stated that when Disability Services and the College of Liberal Arts developed their policy on foreign language course substitutions, they did exactly what Judge Saris stated was done by Boston University.

Randall Borst speculated that it is the social phenomenon that makes the case so interesting, not the legal one. He added that college and university presidents hold a great deal of power and it is necessary for them to be convinced that people with disabilities are educationally and financially worth accommodating. Ron Blosser (University of California at Irvine) felt that the interest generated from the BU lawsuit stems from the fact that Boston University's curriculum and foreign language requirement is pretty much the norm for that type of institution. He feels that this similarity with many other campuses is the reason why Boston University's case has such relevance to others and is being so closely followed.

In an analysis of the final decision, Jane Jarrow (DAIS) felt that the institutionality of the requirements was the basis of the decision. She maintains that although it was determined that BU had an institutional requirement for foreign language, there are very few institutions out there that do the same. She declared that unless the institution has these broad curricular requirements for all students, then the BU case has little relevance to other schools. In Jarrow's opinion, most schools that grant course substitutions will continue to do so, as will those schools that wish to maintain or establish more rigid degree requirements. In the end, she agreed with the decision by Judge Saris in the BU court case.

What continues to be interesting in this medium for sharing and exchange of information, is the many different angles the BU case has been looked at. Russ Pollak (Manhattan College) focused his reaction on the importance of making the "personality," as he called it, of the DSS ofice compatible with that of the host institution. He wrote that the cornerstone of success is to have the DSS office and the institution constantly evaluate and adjust their stances and be attuned to one another's wants, desires, and needs. Adding to Ross Pollak's point, Stephanie Bell (Cabrini College) stated that Westling did not make any effort to adjust his posture to the goals, efforts, and obligations of the LDSS office at BU, an office of long standing and high regard. Stephanie Bell's main concern was that it seemed to her that the final decision ignored everything written over the years about a need for case-by-case decisions.

The determination of which aspects of the curriculum are fundamental are often made in advance of any request for a particular accommodation, and according to Michael Masinter (Nova Southeastern University), when this is the case, no individual assessment is necessary since no accommodation is necessary. In short, he states that the determination of which aspects of a program are fundamental can be made without regard to any particular disability and, therefore, can be made without consideration of the needs

of a particular student. Jarrow (DAIS) supported the judge's decision as well, writing that since BU's requirements were based on sound academic judgment, they are not required to test that particular decision on a case-by-case basis.

Sam Goodin's (University of Michigan) concern centered on what will happen next. He felt that regardless of the decision or BU's stance on foreign language requirements, when presented with the right kind of student in the future (the examples given were a valedictorian, class president, or Oxford scholar), BU would indeed declare them to have mastered the language when indeed they have not to allow them to obtain their degree. An agreement to Sam Goodin's opinion was posted by Opperman (KippOpp Consulting), who felt that institutions bend the rules when it is in their best interests to do so, usually meaning that money is involved.