Technology in Education
Virginia Society for VSTE
www.vste.org

Continuous Speech Recognition Technology: Educational Applications and Best Practices

by Richard C. Snider, Ph.D.

There are a growing number of articles exploring the effectiveness of speech recognition technology in helping students with learning disabilities compensate for written language difficulties. Despite the potential advantages of this technology in supporting students with learning disabilities, little effort has been made to look at the software and suggest best practices for the utilization of speech recognition in the classroom.

The purpose of this article is to introduce speech recognition technology, provide a brief overview of the literature involving speech recognition and individuals with learning disabilities, and then offer several practical suggestions for getting started with speech recognition technology in the classroom.

Introduction to Speech Recognition Technology

Speech recognition is the ability of a computer and program to recognize and carry out voice commands or take dictation. Using speech recognition software, users can tell computers to execute commands and dictate text directly into a word processor on a computer.

In general, speech recognition software used for dictation involves the process of the user speaking into a microphone, the computer processing the spoken words through a sound card, the software analyzing the sounds and matching them against a template, and the matched words appearing as text in a word processor (Raskind, 1993). Some modern speech recognition systems can be used immediately, but most actually learn the characteristics of each person's voice over time, resulting in increased accuracy the more the system is used.

Discrete Speech

Throughout the early to mid-1990s, all marketed speech recognition systems relied on discrete speech technology. Using this technology, users were required to pause between words during dictation. These systems were classified as speaker-dependent, which means that each user had to train the system to recognize his or her dictated speech. The training consisted of users reading selected text passages and took anywhere from one to three hours to complete (Cavalier & Ferretti, 1996).

Speech Recognition, continued

Continuous Speech

In April 1997, Dragon Systems, Inc. marketed the first continuous speech recognition system, Dragon NaturallySpeaking. This system allowed users to dictate text into the computer using natural conversational speech (Dragon Systems, 1999). Unfortunately, the use of the more complex continuous speech technology required much more memory than the previous discrete speech software, and if a computer did not have the necessary memory to run the software, accuracy was greatly reduced (Williams, 1998).

Like the speaker-dependent discrete systems, the NaturallySpeaking user would train the system by reading from a set of text passages, which allowed the software to closely match his or her dictation style. The more consistently the user dictated the text passages, the better the system's recognition rate (De La Paz, 1999).

Current continuous speech recognition systems, such as IBM's Via Voice and ScanSoft's NaturallySpeaking, allow the user to dictate text into the computer in a more conversational manner, but they still aren't completely natural because accuracy depends on consistent and clear pronunciation. Although most systems capitalize the first word of each sentence, in order to dictate effectively, the user must also learn and use commands for capitalization, punctuation, and modification of dictated text. Most systems allow the user to format and edit dictated text using either voice commands or the keyboard (De La Paz, 1999).

Even though manufacturers claim accuracy rates of 95-99% (Essex, 1999), many factors affect the actual accuracy of these systems. These factors include what content is being dictated, how similar the words or phrases are to one another and the variability in the user's speech that may occur due to fatigue, a cold, or mispronunciations. Other factors that can influence accuracy include the environment in which the system is used, such as placement of the microphone and the surrounding noise level, and the quality of the equipment used with the system (Williams, 1998).

Once the user becomes accustomed to using continuous speech input and the software has been trained to recognize his or her voice, it is possible to achieve input rates of up to 130 words per minute. Additionally, most modern products provide support for commonly used packages like Microsoft Word, so it is not necessary to learn how to use a new set of word processing features in addition to becoming familiar with speech input commands (Williams, 1998).

Applications

Continuous speech recognition technology has the potential to make a great difference as a solution for individuals with disabilities. With current speech recognition systems being relatively easy to use, cost efficient, and capable of running on a standard computer system, many researchers and teachers are beginning to explore the use of this technology to assist individuals with disabilities in school, at home, and in the workplace. Examples include the use of speech recognition systems to control the environment for individuals who have physical and/or cognitive disabilities and the use of the technology to help improve the speech accuracy of individuals with hearing impairments (Cavalier & Ferretti, 1996).

Speech Recognition, continued

Recently, a few researchers have begun investigating the use of speech technology as an alternative for students with learning disabilities to get their thoughts down on paper (Higgins & Raskind, 1995; Raskind & Higgins, 1998; De La Paz, 1999).

Learning Disabilities and Speech Recognition

Higgins and Raskind (1995) conducted an experimental study indicating that speech recognition is beneficial to postsecondary students with learning disabilities in the area of written composition. The researchers indicated that discrete speech recognition promotes use of the more developed oral vocabularies of the participants as indicated by the use of larger words.

Raskind and Higgins (1998) followed their 1995 study with a three-year longitudinal study that looked at academic, behavioral, and attitudinal changes as a result of using speech recognition. The data were collected using interviews, questionnaires and self-reports. Over the three-year period, participants significantly increased their GPAs and their use of the lab-based speech recognition systems. The overall attrition rate of the students also decreased during that time. Furthermore, an examination of databases documenting use of services and data from several questionnaire responses indicated that students who participated in the study increased their overall independence by relying less on family members, friends, and classmates to help them compensate for their disabilities.

Although there are currently no published research articles to support the assertion that continuous speech recognition technology can offer students with learning disabilities in the area of written expression a superior method to write as compared to traditional methods, several authors have cited this technology as a possible tool to support these students. MacArthur (1999) and De La Paz (1999) both indicated that dictation using continuous speech recognition software has an advantage over keyboarding and handwriting because it helps circumvent the issues of producing words. The authors also mention that speech software offers individuals with learning disabilities the possibility of composing by dictation without having to rely on another person to transcribe their writing. De La Paz, however, does mention that there are potential difficulties with using continuous speech recognition for individuals who have learning disabilities. The author indicated that although continuous speech recognition, like dictation, may help users circumvent the production difficulties of writing, it creates additional demands that include careful speech, error correction and editing procedures. In order to address these issues, De La Paz recommended using knowledgeable educators to guide individuals with learning disabilities as they begin using the technology.

Suggestions for Using Continuous Speech Recognition in the Classroom

Although the literature seems to indicate that continuous speech recognition technology can be potentially useful to assist students with learning disabilities to get their thoughts down on paper, there could be difficulties for the classroom teacher in using this technology if it is not properly implemented. The following recommendations are based on my most recent reviews of the literature and my experiences in using speech technology for over eight years as a K-12 special education consultant and as a university researcher. These suggestions are not

Speech Recognition, continued

specific to any one software package, but can be generalized to any number of speech recognition programs.

Guided Training

Guided training sessions are invaluable in teaching students to use speech recognition technology. When properly implemented, guided training sessions allow educators to teach students techniques for correctly using the software and allow them to monitor student progress and provide immediate feedback to help improve their skills. This type of training session is likely to assist students with disabilities to overcome initial frustration with using the speech recognition technology and allow them to get comfortable with using the software. Guided training is important as the students learn to create their voice models, dictate, and edit their work.

Creating the Voice Model

Although some speech systems will allow users to begin dictating without first training the software to recognize their voices, these systems are typically less accurate until the software adapts to the user's voice. Most speech systems will require the user to read several selected stories first in order for the software to adjust the factory template to the user's voice enough to accurately recognize dictated text. Typical training times are usually about thirty to forty-five minutes.

Usually the systems will have two to four stories and may require the user to read half of them in order to begin dictating. Since it tends to increase the overall accuracy of the speech software, it is beneficial to have the students read all of the available training text before they begin dictating for the first time. Remember, however, this does not have to be done in one session and, in fact, should probably not be since the student will tire. Break up the training into several sessions and let the students take breaks during each session so they will be at their best as they read the text. Allowing the students to get a drink of water is also helpful after reading long sets of passages. If a student gets tired or uncomfortable during training, he/she is more likely to make errors during reading and this will reduce the accuracy of the speech software. This type of voice training will maximize the initial accuracy of the software and minimize frustration as the students begin using the speech recognition technology. This may increase the training time, but it is well worth it for the students.

Unfortunately, students with learning disabilities having more severe writing and accompanying reading disabilities often have difficulty reading the stories required for training the speech technology systems. In this instance, depending on the software and the student, it may work to turn off the microphone, read a sentence of the training story to the student, turn the microphone back on, and have the student repeat the sentence. This may take a little practice and may require more time to complete the training, but could allow the student ultimately to use the software.

Dictation

Once the student has trained the software and begins dictating into the word processor, it is necessary to monitor progress and provide feedback. As a knowledgeable educator, it is important to know that the continuous speech

Speech Recognition, continued

systems use algorithms that work best when the users speak phrases, sentences, or even paragraphs at a time – not single words. Although many continuous speech systems claim a user can speak in a "conversational style," it is vital for the user to slow down his or her speech and to enunciate his or her words a little more clearly than normal during speech. Often when students first begin to dictate, they will mumble, which results in a high error rate, but when reminded to slow down and speak clearly the students are able to achieve a much higher accuracy rate while using the speech software.

Editing

Individuals with learning disabilities may have more difficulty with the editing process using continuous speech recognition technology than they do with typical discrete speech systems. Higgins and Raskind (2000) noted in a study on remedial effects of speech recognition that correcting individual words using the continuous speech systems may be more difficult since these words may be harder to identify and isolate within phrases or sentences, whereas the older discrete speech systems allowed users to dictate and correct words one at a time.

Using the voice to get the text into the computer is the primary advantage of voice recognition for individuals with learning disabilities. Allowing students to use the mouse and keyboard to make corrections to the dictated text reduces the cognitive demands of trying to do so by voice and helps with the editing process. Finally, although there is little or no research to directly support this assertion, multi-modal communication theory would support using computer-based voice output technologies to highlight and read the dictated text to the students in order to help them edit the content.

Conclusions

Continuous speech recognition technology has the potential to work well for individuals with learning disabilities to get text into a word processor if implemented properly through the use of guided training, thorough training of the voice model, and clear and precise dictation of the text. This technology can be combined with the mouse and even the keyboard for editing of the text if necessary. As an educator, do not be hesitant to use continuous speech technology with other tools if it will help you accomplish your primary goal – to provide a better learning environment for your student.

References

- Cavalier, A. R., & Ferretti, R. P. (1996). Talking instead of typing: Alternate access to computers via speech recognition technology. Focus on Autism & Other Developmental Disabilities, 11(2), 79-85.

- De La Paz, S. (1999). Composing via dictation and speech recognition systems: Compensatory technology for students with learning disabilities. Learning Disability Quarterly, 22, 173-182.

Speech Recognition, continued

References

- Dragon Systems, Inc. (1999). About the company. Available: http://www.dragonsys.com/about/index.html [1999, June 20].
- Essex, D. (1999). Continuous speech recognition. Healthcare Informatics. Available: http://www.healthcare-informatics.com/issues/1999/02_99/nine.htm [1999, November 20].
- Fishman, E. S. (1996). History and development of voice activated software. 21st Century Eloquence. Available: http://www.voicerecognition.com/physicians/ medical_office.html [1999, November 18].
- 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(2), 159-174.
- Higgins, E. L., & Raskind, M. H. (2000). Speaking to read: The effects of continuous vs. discrete speech recognition systems on the reading and spelling of children with learning disabilities. Journal of Special Education Technology, 1(15).
- MacArthur, C. A. (1999). Overcoming barriers to writing: Computer support for basic writing skills. Reading and Writing Quarterly, 15, 169-192.
- Raskind, M. (1993). Assistive technology and adults with learning disabilities: A blueprint for exploration and advancement. Learning Disability Quarterly, 16(3), 185-196.
- Raskind, M. H., & Higgins, E. L. (1998). Assistive technology for postsecondary students with learning disabilities: An overview. Journal of Learning Disabilities, 31(1), 27-40.
- Williams, M. (1998). Guide to speech recognition. PC Magazine. Available: http://www.zdnet.co.uk/pcmag/supp/1998/speech/.

About the Author

Richard Snider is currently Technology Coordinator at the Training and Technical Assistance Center (T/TAC) at Virginia Polytechnic Institute and State University, Blacksburg, Va. He can be reached at: rsnider@vt.edu.