# UTILIZING SPEECH RECOGNITION TECHNOLOGY TO INCREASE PRODUCTIVITY -- CASE STUDIES

Hong Z. Tan & David M. Horowitz
Center for Rehabilitation Technology
Dept. of Rehabilitation Medicine
Tufts University School of Medicine, Boston, MA

## ABSTRACT

A comprehensive vocational service delivery model using speech recognition technology was presented earlier [1] and has been applied with success. Here, we present two case studies. One of the two clients, JEM, is an architectural designer, who is quadraplegic, and who wishes to regain self-employment. The other client, HK, is a vocational rehabilitation counselor, who suffers from multiple sclerosis, and who requires assistive technology to increase her productivity on the job. While HK's main concern is to document her cases in a timely fashion (she is barely able to type), JEM needs to document building specs as well as to draft blueprints. JEM's rehabilitation program involves more technology compared to HK's. The results are encouraging. We are convinced that when knowledge engineering[1] is applied to voice recognition technology, productivity is increased. A person with a disability utilizing such assistive technology is in a strong competitive position with his/her able-bodied counterparts. Issues concerning psychosocial/medical aspects of our rehabilitation program are also addressed.

## BACKGROUND

There are approximately 650,000 individuals in the United States who experience complete paralysis of all extremities due to a variety of neuromuscular disorders (source from Washington D.C., Data on Disability from the National Health Interview Survey 1983-1985). The extent to which disability impairs manual dexterity contributes to the problems individuals with quadriplegia face in achieving gainful employment. In view of these statistics, there are many individuals who could benefit from speech recognition technology.

Speech recognition allows individuals to speak to a computer to dictate correspondence or operate software, instead of typing. Despite the enormous potential speech recognition offers individuals with severe disabilities, there are very few detailed case studies in the rehabilitation literature on the vocational applications of speech recognition. (For a more thorough literature review, see [1]).

Although there has been discussion of the benefits of speech recognition technology to serve the vocational needs of persons with severe disabilities, the technology has not begun to serve all those who could potentially benefit from it. With the introduction of large vocabulary speech recognition technology in 1985, rehabilitation professionals anticipated very optimistic results on the number of individuals who would return to work [2]. Although many individuals acquired or purchased expensive technology, only a small percentage of those individuals who received speech recognizers are currently employed. Part of the problem lies with the fact that the technology has not been ready for distribution to individuals who are quadriplegic. In part, it has been impossible to predict based on client profiles, which individuals stand the best chance of vocationally succeeding with the technology. This problem is due to the fact that little systematic or thorough clinical reporting has been published in the rehabilitation literature.

Another important part of the problem is the fact that speech recognition is slow. The focus has been that individuals who are disabled are now able to do something they were not able to do before through the use of speech recognition. However, consider the employer's perspective. The employer wants a cost-effective and productive employee. Computer based jobs involving documentation or software operation can be objectively and quantitatively evaluated in terms of an individual's productivity. One parameter the employer needs to consider is text creation rate. The job the employee holds, whether it be computer programmer, receptionist, administrator or social service counselor, is required to create text to complete the job.

Large vocabulary speech recognition technology is a severely rate limited way to create text. Few papers are reported in the literature which systematically study and report text creation rates with current speech recognition technology. Antectdotal reports claim text creation rates of 15 to 40 words per minute, depending on the nature of text being created; creating new text takes longer than creating memorized text. Comparing the rates of users of large vocabulary speech recognition to able-bodied typists makes it difficult to argue that speech recognition technology permits individuals to produce on the job in a competitive manner. At the high range of the scale, able-bodied typists are able to create text at approximately 100 words per minute. Court recorders who use a special method of keyboarding are able to type at approximately 150 words per minute while conversational speech ranges anywhere between 150 to 300 words per minute [3]. If the employer is under the illusion that speech recognition will be fast because speech is fast, this may result in an initial enthusiasm for the prospect of hiring an individual with severe writing impairments. However, without appropriate counseling from a rehabilitation service delivery team, the employer may become disappointed after comparing the productivity results of the speech recognition user to the keyboard user.

There are instances where users report that use of a speech recognizer results in faster documentation than traditional methods of text creation. Emergency medicine physicians, radiologists and pathologists report a dramatic decrease in the time it takes to generate a written report by making use of special applications of speech recognition technology [4]. Here, the speech recognizer is used to access pre-stored blocks of text. Often referred to as voice-macros, these macros limit the amount of text that needs to be explicitly spoken, thereby increasing the rate of text creation. While creation of a large number of voice-macros (which can exceed 5,000 macros) is a time-consuming process, the resultant boon in productivity can be well worth the effort. Adopting the approach of job specific application development, we predict the potential of speech recognition in vocational rehabilitation can best be achieved within an appropriate clinical framework [1]. This paper focuses on one technology which permits job specific application development. Its utilization is illustrated through two case studies.

[1] We use the term "knowledge engineering" in the sense that voice macros are created for highly repeatable blocks of texts, and these macros are hierarchically structured.

The speech recognition technology used in this model is the Kurzweil Voice Report system (the KVR) manufactured by Kurzweil Applied Intelligence Inc. (KAI, Waltham, Massachusetts). The KVR is built on three layers: (1) the large vocabulary voice recognition system that has functionally unlimited vocabulary size, is speaker independent/adaptive, uses discrete speech, and responds in real time; (2) the report generation software that takes care of the format of the reports the KVR generates; and (3) the domain-specific knowledge base that contains domain-specific vocabularies, and uses trigger phrases and fill-ins to achieve increased productivity with structure and flexibility. The trigger phrase is a word or a phrase you say to bring up a predefined block of text. This block of text might contain fill-in-the-blanks, or so-called "fill-ins". Trigger phrases allow the user to say very few words and yet produce large chunks of texts; "fill-ins" keep the trigger-to-text translations flexible. A knowledge base is a set of trigger phrases about a particular topic (i.e., domain).

## OBJECTIVE

The overall goal of our speech recognition rehabilitation program is to identify clients whose (potential) job will benefit from the utilization of speech recognition technology, to develop speech recognition applications specific to the job, to train the client, and to deliver the technology to the client so s/he can be gainfully employed.

Our first client, JEM, is a 50 year old architectural designer who was self-employed prior to his spinal cord injury in 1988. A C5 injury left him quadraplegic. Due to the sensory and motor impairments in his hands, he is no longer able to fill out detailed specification sheets, or to draft plans for a house; nor can he use a conventional keyboard. Before JEM was enrolled in our speech recognition rehabilitation program, he used a double arm sling to aid his writing. However, this resulted in illegible writing due to hand tremor. Furthermore, he could not do any precise drawing with the arm sling. Since JEM's speech is clear (although somewhat slow due to vocal cord injury), he was enrolled in our speech recognition rehabilitation program. JEM's rehabilitation goal is to be self-employed again with the aid of assistive technology.

Our second client, HK, is a 43 year old vocational rehabilitation counselor with the state vocational rehabilitation agency, who was diagnosed with multiple sclerosis in 1979. She can barely hit the keyboard with her left hand, and she can't control her right hand due to intention tremor. Her rehabilitation goal is to be able to document client meetings and phone conversations in a timely fashion using her voice as computer input so that she can retain employment.

## APPROACH

The approach we take for each client is somewhat different depending on the individual's disability and the job requirements. For JEM to be gainfully self-employed as an architectural designer, he has to produce (1) detailed written documents (i.e., contracts, specification sheets, material lists, etc.); and (2) architectural drawings with varying degrees of detail (i.e., floor plans, cross sections, etc.). JEM's rehabilitation program involves three phases: (1) using voice recognition technology for document production; (2) using voice recognition technology together with other means of input devices (e.g., head-mounted pointer) to operate CAD software to produce architectural drawings [5]; and (3) system integration. We've finished the work for the first phase and the results will be reported in the next section. The second phase of the project is well under way. Our plan for the second and third phases of the project will be presented in the discussion section.

HK's rehabilitation program is essentially the same as the first phase in JEM's program.

## RESULTS

All our training on the KVR was done in a noisy office environment, where people come and go and have meetings. The recognition rates were greater than 90% for both JEM and HK. We believe that this high recognition rate can be maintained when JEM uses his KVR at home or when HK uses hers in her office.

JEM has received over 100 hours of training. The domain-specific knowledge base developed for him allows him to produce documents. The KVR allows JEM to produce professional-looking documents efficiently. It takes JEM 7 minutes to complete a 2-page specification sheet and print it out on a laser printer. According to JEM, it used to take him 40 minutes to fill out the same spec sheet by hand prior to his spinal cord injury. Post trauma, it would take him up to 2 hours to finish the same work and his handwriting was only legible to himself. We estimate that it will take JEM approximately 2 hours to dictate and format the same form word by word using a conventional large vocabulary isolated word recognition system.

Due to his vocal cord injury, JEM's speech is slow and changes over time. We have observed that after 2 or 3 hours of continuous speaking, JEM's speech is even slower[2]. His intonation changes as he struggles to utter words. With the voice-macros we've developed for JEM, he can produce documents well within the 2-hour limit during which he is in good voice.

HK has received roughly 20 hours of training. The KVR has been demonstrated to benefit HK. In addition to being able to access a computer through voice and to produce documents very efficiently with trigger phrases, HK also benefits from cue text[3]. HK is required to put codes next to almost everything she documents. While she remembers some of the most commonly used codes, she has to look up other codes because HK also suffers from memory deficit. Given the mobility impairments with her hands, this is both cumbersome and time-consuming. With the KVR, we have incorporated code lists in cue text, so HK can see the codes next to their meaning, and be able to select one without going through her files.

## DISCUSSION

We've began phase two of JEM's program which involves: (1) training JEM to use AUTOCAD software, and (2) developing necessary software so JEM can access AUTOCAD through voice and/or head-mounted pointer [5]. Phase three of JEM's program is system software/hardware integration. We first became aware of this issue when we moved from phase one of the project into phase two. At that time, the KVR software only ran on a 386-PC while AUTOCAD ran on both the 386- and 486-PC. The question was what type of PC should we recommend to JEM. We later found out that the KVR system

---

2 It was objectively and quantitatively determined to take him twice as long to pronounce a word compared to a person without vocal cord injury.

3 Cue text is associated with "fill-ins". It is displayed on computer screen during dictation of documents, but does not enter the final print-out of documents.

also runs on a 486-PC. The 486-PC was chosen. In JEM's case, more sophisticated software and hardware are needed in order for him to perform at a competitive level. We need to be aware of software and hardware compatibility issues from the beginning. In planning for system integration, we also asked JEM to draw a floor plan of his office/drafting area in his house (JEM will be working at home), so we can begin to plan how to position equipment into his work space, and whether further modifications to his house will be required.

For HK, we will continue to implement additional forms. We have finished work on the "initial interview form", and have begun working on "certificate of eligibility", "fiscal contact report", and other forms.

Reliable large vocabulary voice recognition technology is still new and very expensive (>$8,000). We chose the KVR system because it increases productivity dramatically through knowledge engineering. We believe that in order for a person with a disability to be gainfully employed, s/he needs to perform competitively compared to other people in the work place. As long as the technology is well supported and has potential for further development, it should be demonstrated to be cost-effective to use.

The issue of maintenance and customer support arises whenever high tech is involved. Given the sophistication of the KVR and AUTOCAD software, JEM can not be expected to solve all future problems as he might encounter. Thus it is essential to have an adequate plan for his continuous support long after the technology has been installed and after his case is closed in our program. We have worked closely with customer support at KAI and have reason to believe that their customer support team will adequately support JEM. During the second phase of our program, we need not only to train JEM on CAD, but also to establish adequate long-term technical support for him on this technology as well.

Psychosocial and medical aspects

In a rehabilitation program that involves high tech training, the engineer deals with the client most of the time. However, the engineer might not be aware of all the medical and psychosocial aspects about the client which we believe are important to the success of technical training as well as the client's future success. Our program includes on its team engineers, a clinical/research psychologist, a speech scientist, and a physiatrist, who work together to promote the well-being and success of our clients.

Before JEM started his training on the KVR system, the engineer was aware of his short-term memory loss and was able to employ strategies and utilize features in the KVR to provide cues to guide him through dictation. An intensive training was done right after JEM was introduced all the basic features of the KVR. It proved to be quite effective.

As the training went on, it was noticed that JEM often went astray during conversation and would go on a tangent and not return to the topic. JEM appeared to be manipulative in negotiating equipment needs. The psychologist discovered that JEM's lack of focus was due to minor brain injury. Therefore, he was not able to suppress random thoughts in the midst of conversation.

It was also observed that JEM seems to feel that the more equipment we recommend, the more we support him. He fears losing the emotional or technical support from our program. When we recommended that the computer monitor and the monitor for his VisualTek device be combined so only one monitor needs to be purchased, JEM insisted that he get two monitors. JEM apparently has difficulty letting go any equipment he wants to get. This will be negotiated with JEM's state VR counselor and JEM to suit both party's needs.

When the engineer who worked with JEM for the first phase of the project noticed something unusual when the phase came to a close, our clinical psychologist was able to point out the difficulties some clients have with termination of any kind. Thus it was made clear to JEM that he will receive continuous reinforcement training from the engineer as he moves into the second phase of the project. This helped JEM focus on the work again.

We have spent much less time working with HK and are only beginning to discover her psychosocial and medical aspects. HK is currently working and uses a computer at work site. She fears that learning to use the KVR at the same time will cause confusion on her job (due to her cognitive limitations). We'll need to evaluate her concern and when necessary, work out strategies to help her better cope.

JEM and HK are both highly motivated individuals who expect our speech recognition rehabilitation program to contribute to their gainful employment. Productivity is key to their future success. With the KVR system and knowledge engineering, we have been able to provide both clients with a document production tool that allows them to produce professional reports in far less time than with conventional speech recognition systems.

## ACKNOWLEDGEMENTS

This work was supported by the United States Department of Education Rehabilitation Services Administration Grant #H128A01011-90. Special thanks to Betty Sniegoski and Kathy Rowe for their help on numerous aspects of this work.

## REFERENCES

[1] Horowitz, D. and Lash, M. (1991). "A comprehensive rehabilitation technology service delivery model utilizing speech recognition technology". Proceedings of the 14th Annual Conference of RESNA, Kansas City, Mo.

[2] Shane, H. (1991). "Special session of Voice I/O for individuals with disabilities". Presentation at the 14th Annual Conference of RESNA, Kansas City, Mo.

[3] Nickerson, R.S., Huggins, A.W.F., and Stevens, K.N. (1978). "Some observations on the temporal properties of speech", Bolt, Beranek and Newman, Inc., Report No. 40, Cambridge, Ma.

[4] Kurzweil, R. (1992). Personal communication.

[5] Burnett, J., Klabunde, C., and Britell, C. (1991) "Voice and head-pointer operated electronics computer assisted design workstations for individuals with severe upper extremity impairments", Proceedings of the 14th Annual Conference of RESNA, Kansas City, Mo.

Address correspondence to:
Hong Z. Tan
Rehabilitation Medicine
75 Kneeland Street 5th floor
Boston, MA 02111