A COMPREHENSIVE SPEECH RECOGNITION SERVICE DELIVERY MODEL ILLUSTRATED WITH CASE STUDIES

Hong Z. Tan, Robert R. Ebert, David M. Horowitz, Corine A. Bickley
Center for Rehabilitation Technology
Department of Rehabilitation Medicine
Tufts University School of Medicine, Boston, MA

ABSTRACT

A comprehensive vocational service delivery model using speech recognition technology has been applied with success. The model brings to bear important technological and industrial advances in speech recognition technology which permits computer access for individuals who can not use a keyboard due to neuromuscular impairments. The model goes well beyond the traditional approaches to applying speech recognition to rehabilitation which often results in the purchase of the device but not employment. It brings together the skills of a university-based rehabilitation engineering program and a state vocational rehabilitation commission. The collaborative effort results in a program which matches the strengths of clients with the employment opportunities in the state. Once the client has identified a career plan, knowledge engineering[1] specific to the job is applied to maximize the client's productivity. Two case studies will be presented which demonstrate that when job specific applications of speech recognition are used, the technology enhances productivity more so than traditional uses of speech recognition which involve word by word dictation of unstructured text. A person with a disability utilizing such assistive technology is in a strong competitive position with his/her able-bodied counterparts.

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. Thus, 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 using a keyboard. 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 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 only begun to serve some of those who could potentially benefit from it. With the introduction of large vocabulary speech recognition technology in 1985, rehabilitation professionals anticipated very optimistic results with regard to the number of individuals who would return to work [2]. Although many individuals acquired 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 addition, 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, usually requires the production of substantial mounts of text.

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. Anecdotal 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

[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.

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 that 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.

THE KURZWEIL VOICE REPORT SYSTEM

The speech recognition technology used in our project is the Kurzweil Voice Report system (the KVR) manufactured by Kurzweil Applied Intelligence Inc. (KAI Inc., 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 that brings up a predefined block of text. This block of text often contains blanks to be filled in ("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 and the associated predefined blocks of texts structured hierarchically.

CASE STUDIES

Case Study 1

Our first client, HK, is a 43 years old vocational rehabilitation counselor with the Massachusetts Rehabilitation Commission. She 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 (i.e., her right hand tremors whenever she intends to use it). 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.

HK's rehabilitation program is essentially using voice recognition technology for document production. We have started implementing the "MRC Initial Interview" form for her. We plan to implement more forms, e.g. eligibility evaluation, for her in the future. She has received over 20 hours of training and 50-100 hours of more work is expected.

We're beginning to see how the voice recognition system can help expedite HK's document production. For instance, HK is required to put codes on almost everything she documents. Given the mobility impairments with her hands, it is both cumbersome and time-consuming for her to look up codes in a file or a book. because she also suffers from a memory deficit, she can only remember a few most frequently used codes. We have implemented several code lists, including codes denoting each type of disability, into her knowledge base. With this system she is shown codes and their meanings on-screen so she can select one from the list. By selecting a code vocally, the code as well as the meaning of the code is put into the document and is ready to be printed out with another voice command.

It is anticipated that this application will be utilized by a broad population of vocational rehabilitation counselors within the nationwide RSA network of vocational rehabilitation agencies, independent of whether the counselor is disabled or not. The anticipated gain in productivity is expected to be attractive to all rehabilitation counselors.

Case Study 2

Our second client, JEM, is a 50 years old architectural designer who was self-employed prior to his spinal cord injury in 1988. A C5 injury left him quadriplegic. 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 9 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. The arm sling also caused upper extremity muscle tension and fatigue which made it impossible for it to be used for a prolonged period of time (15 to 20 minutes maximum). Furthermore, he could not do any precise drawing with the arm sling. Since JEM's speech is clear (although somewhat slow and deteriorating after 2 hours 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

For JEM to be gainfully self-employed again 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 speech recognition technology for document production; (2) ,using voice recognition technology together with other means of input devices (e.g., head-mounted pointer, an eyetracker, etc.) to operate CAD software to produce architectural drawings [5]; and (3) system hardware/software integration.

JEM has finished the first phase of his rehabilitation program. We have implemented several catalogs into his knowledge base, which include, but are not limited to, window catalogs by several leading manufacturers. JEM has received over 100 hours of training on the domain-specific knowledge base developed specifically for him. The KVR system allows JEM to produce professional-looking documents efficiently. For instance, 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 a '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. As stated earlier, JEM can only talk for 2 hours continuously before his speech quality deteriorates due to injury to his vocal cords. With the knowledge base we've developed for him, he can produce a spec sheet, a material list, and a contract well within the 2-hour limit during which he is in good voice. Therefore, the effort of knowledge engineering has proved to be eminently worthwhile.

The second phase of JEM's rehabilitation program is well under way. We've begun the process of training JEM to use AutoCAD and to evaluate the peripheral equipments which JEM will need to interact with AutoCAD. These pieces of equipments will provide voice and head-pointing control of selection and positioning which is standardly accomplished through keyboard input and mouse positioning. JEM is not able to use a keyboard or mouse for any extended period of time. The AutoCAD system facilitates the production of high-quality drawings. Our plan is to develop a set of drawings, which represents the basic house plans which JEM has used in the past in his semi-custom design of houses. His design process is semi-custom because he has usually worked with a group of standard house plans which he then customizes based on discussions with his client. The basic plans will be implemented in AutoCAD as part of the second phase of JEM's rehabilitation program. JEM will then be able to use these drawings to create customized house plans immediately after completion of the third phase, system integration in JEM's office. JEM is currently in the initial phases of AutoCAD training. We are in parallel assessing JEM's ability to see the AutoCAD positioning crosshairs on a background color; both the colors of the background and of the crosshairs can be varied. After JEM completes training on AutoCAD, he will begin voice training for AutoCAD keyword commands. He will then practice using the HeadMaster for AutoCAD pointing functions.

The third phase of JEM's rehabilitation program, i.e. system hardware/software integration, is necessary because JEM needs sophisticated software (i.e., the KVR, the AutoCAD, etc.) and hardware (i.e., 486 PC, laser printer, paper cutter, etc.) in order 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 that we can begin to plan how to effectively position equipment in his work space, and determine whether further modifications to his house will be required to enhance efficiency..

Given the sophistication of the KVR and AUTOCAD software, JEM cannot be expected to solve all future problems that he might encounter. Thus it is essential to have an adequate plan for continued support long after the technology has been installed and his case closed in our program. We have worked closely with customer support at KAI and have reasons 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.

Medical and Psychosocial Issues

The purpose of this section is to explore several problems that arose in helping JEM but did not arise with other clients who have been worked with. As have been described, JEM has worked hard, learned much and made substantial progress toward being able to run his own business. The problems that arose should allow us to consider some important issues in running an effective rehabilitation team.

During the successful progress of JEM's training, the engineering/managerial staff began to perceive problems in aspects of their relations with JEM. JEM appeared to often go astray during conversations, going off on a tangent without returning to the topic. This was interpreted at times as purposeful misdirection. He was also seen as manipulative in negotiating equipment needs-seeming to the staff to equate more equipment with a better and more appropriate equipment system; calling up financial offices at the Tufts medical center frequently and thus annoying the bureaucracy; and also similarly annoying the person in charge of equipment funds at the funding agency. As JEM was now being viewed increasingly with anger as a source of considerable trouble, the psychologist who had just joined the team was asked to intervene both by talking to JEM individually and by participating in a group meeting with JEM to discuss equipment needs. As there were no available records dealing with the history of his disability, nor with any psychosocial aspects of his life, some basic information also had to be investigated.

When asked about how the program was going, JEM was so extremely positive while giving impressions of fear (coupled with the ongoing problems regarding dependence and independence) that the psychologist decided to investigate further, knowing the frequently ambivalent nature of relations between disabled individuals and helping professionals. As the project engineer working closely with JEM has excellent interpersonal skills, it seemed unlikely that the fear was related to the

immediate situation. Indeed, when discussing the accident which left him quadriplegic, JEM stated that the final damage to his spinal cord occurred not at the accident but after hospitalization and so, the quadriplegia was attributed to the treatment received. Whether objectively accurate or not, this view seems to have negatively affected his relations with helping professionals and rehabilitation organizations, even those who he feels are truly helpful. The high degree of anxiety that may then be associated with questions of the adequacy of help received can easily he taken as a lack of trust, veiled criticism or an affront by those trying to help while balancing other agendas such as cost.

With regard to the tendency to go astray during conversations, JEM admitted to this and related it to neurological damage sustained. He stated that during a conversation, an unrelated thought will appear and that he cannot assess its degree of relevance. He proceeds with the new thought while others wait for him to tie it back to the topic at hand. As he does not know it is unrelated to the topic, he just continues on, confusing and distressing those around him. He reported having discussed this with his neurologist at great length and having developed strategies to cope with it whose efficacy was diminished by problems getting to the rehabilitation center (having to rise at 4 am to catch a ride to a three hour bus trip to Boston). This neurologically based problem seems to have been a major source of communications difficulty at the program and could jeopardize future work if the coping strategies being utilized are inadequate. Therefore, direct contact with his current medical team was indicated (none had been made or recorded previously). This is now taking place and a thorough review of coping strategies will be undertaken along with the development of long term support systems around this issue.

Several important points can be made with regard to the problems surrounding the equipment purchase. First, confusing and contradictory messages were given to JEM as to his role in decision making. One team member had told JEM that the team made suggestions and he could decide on the equipment. A second team member felt that JEM should play virtually no part in decision making but should accept what was recommended. This was, however, not communicated directly to JEM. Instead, when he inquired about equipment purchase, the medical center bureaucracy was blamed for delays, thus directing JEM's activities towards prodding various financial offices. It is unclear what he was told about the decision making process by the funding agency. In any case, a certain amount of money has been allocated to him to be spent before a specified date. This money is supposed to be the maximum spent on his equipment, but he does want to see all of it spent and is quite anxious about this. For example, the team recommended that the computer monitor and the VisualTek monitor be combined with a switch to change inputs. This was presented as saving money, but a review of his work space, work flow, and the integration of his hardware/software system had not yet been done (a thorough review of system integration was initiated at this meeting with JEM's involvement). He wanted the two monitors to be separate, assuming (without his own thorough analysis of his workspace) that he would keep the VisualTek separate from the computer. Although the engineering team explained that the combined screen with a switch might be better for JEM, they also repeatedly discussed the monetary savings. He has found another source for partial funding and gone ahead with the purchase of a separate VisualTek, leaving the engineering team frustrated.

GENERAL DISCUSSION

The most useful questions to ask about problems such as we have reviewed do not include "who is to blame?" but are actually "what can be done as a matter of basic policy or preventively to provide for a mutually rewarding and successful rehabilitation process, avoiding potential problems as much as possible?" and "what can be done to resolve problems or disagreements once they have arisen (as some will in most circumstances)?". As seen from the above examples, it became clear that information flow had to tm improved or instituted in a number of essential areas.

First, relevant psychosocial/medical evaluations and contact with outside professionals have to be performed, documented and communicated to the rehabilitation engineering team.

Second, the clients outside medical/psychosocial professional helpers will often need to be integrated into the treatment team.

Third, members of the rehabilitation team need to communicate clearly with each other, with the client and with the agency funding equipment purchases with regard to the structure of decision making in equipment purchase.

Fourth, members of the engineering team need to be aware of the ambivalent relations that can develop with clients (most easily in the absence of clear communications) as well as of the positive and negative changes in dependency/autonomy that can take place because of disability itself; past injuries (physical and/or social or emotional) from professional helpers, family members, and even passers-by (classical stigmatization); hidden agendas in the team (disordered communications); and because of sincere and straightforward feedback about progress towards the mastery of relevant tasks and towards employability.

In terms of the structure of the rehabilitation team, it is essential to view the client as an integral member of the team for almost all purposes. As the purpose of the team is to strengthen the client and make him/her as independent as possible, exceptions to this rule should be considered carefully and reviewed periodically. The degree of technology transfer should be dictated by the interest, background and the abilities of the client as well as the limitations of available time and funding. In terms of decision making, the client should be viewed as one member of the team, each member of the team having an unique and important perspective on the problem at hand. The rehabilitation team dealing with complicated technologies can easily come to believe that the client cannot understand the issues at hand and thus cannot participate knowledgeably in decision making. In fact, complicated technical or scientific issues can usually be condensed to the most important points and presented in plain English (or the relevant language) so that joint decision making can proceed meaningfully. It is, however, the case that many professionals who arc used to conversing in highly technical language (scientists, engineers, physicians, etc.) have difficulty in

translating their thoughts into understandable English propositions. Each team should have at least one such skilled member or access to a similarly skilled consultant. Professionals who perform this type of task regarding medical information are sometimes referred to as health educations and arc present in a wide variety of health settings. Care must be taken that such translations do not reflect any hidden agendas of any team members which would unfairly prejudice and short-circuit the decision making process. During such decision making (for example, the determination of what equipment is finally needed), it should be expected that clients may have concerns as to whether all relevant issues have been considered and their true values properly balanced.

In summary, aside from relevant engineering issues, the rehabilitation team should include the client, be well integrated with the client's outside professional helpers, keep communications clear and well documented, provide for supervision of the engineering staff by an appropriate professional (psychologist, etc.) concerning interactions with clients, and ensure an adequate information flow between all individuals and groups that will affect the client's vocational rehabilitation.

CONCLUSION

JEM and HK are both highly motivated individuals who expect our speech recognition rehabilitation program to contribute to their gainful employment. Productivity is the 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.

REFERENCES

[1] Horowitz, D. and Lash, M. (1991). "A comprehensive rehabilitation technology service delivery model Utilizing speech recognition technology", Proceeding of the 14th Annual Conference of RISNA, Kansas City, Me.

[2] Sham, 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.E, 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, IL (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", Proceeding of the 14th Annual Conference of RESNA, Kansas City, Mo.