Voice Recognition—An Emerging Necessity Within Radiology: Experiences of the Massachusetts General Hospital

Amit Mehta, Keith J. Dreyer, Alan Schweitzer, John Couris, and Daniel Rosenthal

Voice recognition represents a technology that is finally ready for prime time use. As radiology services continue to acquire a larger percentage of the shrinking health-care dollar, decreasing operating costs and improved services will become a necessity. The benefits of voice recognition implementation are significant, as are the challenges. This report will discuss the technology, experiences of a major health-care institution with implementation, and potential benefits for the radiology practice.

Copyright © 1998 by W.B. Saunders Company

ADVANCES IN SPEECH and voice recognition (VR) technology have brought about more practical, useable speech recognition capabilities. In late 1994, the effort to develop large-vocabulary, continuous speech recognition systems in American English had progressed to the point where accuracy had become a less of an issue and vendors could focus on specific markets. In a recent survey, the Healthcare Information Management and Systems Society found in a worldwide membership survey that 31% of respondents were eager to install VR systems in their practices¹; no specialty was more suited than radiology. Until now, radiologists have either handled the process of dictation manually or on tape to medical transcriptionists; the process has been time-consuming and expensive. The choice of transcribing radiology reports with VR systems is an obvious focus due to the consistent and predictable vocabulary within dictation practices. Multiple solutions have been generated and we will discuss our experiences at the Massachusetts General Hospital (MGH) with IBM Medspeak/Radiology (White Plains, NY). We review the VR technology, challenges encountered through deployment, implementation strategies, and experienced benefits.

THE TECHNOLOGY

VR software encompasses, at the highest level, four core technologies: spoken recognition of hu-

man speech, synthesis of human readable characters into speech, speaker identification, and verification. These are referred to as speech recognition (or speech-to-text), speech synthesis (or text-to-speech), speaker identification and verification, and natural language understanding (NLU).

Speech Recognition

Speech recognition, also referred to as speech-to-text, is the technology that makes it possible for the computer to translate spoken word into type. The first type of speech recognition is spoken command recognition, also known as command and control, and the second is pure dictation. Command and control dictation handles the recognition of single words or short phrases spoken with continuous speech, such as "Begin Dictation" or "Accept and Sign." Dictation comes in two divisions: discrete (or isolated) and continuous. The discrete dictation technology has lower processing power requirements, but requires the end-user to place a short pause between each spoken word. Continuous dictation, as the name implies, does not carry this limitation and is used in the radiology implementation.

Speech Synthesis

Speech synthesis, also referred to as text-to-speech, is the technology that makes it possible for the computer to produce the phonemes we would make when we read text aloud. Within radiology, speech synthesis allows the radiologist to review the report, as well as in playback during dictation of reports.

Speaker Identification and Verification

Speaker identification and verification are two related technologies. These technologies, unlike speech recognition and speech synthesis, deal with the identity of the human speaker and not with what was spoken or with synthesizing the particular human voice. With speaker verification, technology is applied to authenticate a given human speaker against a database pool of enrolled candidates.

NLU

NLU is the technology that makes it possible for the computer to understand the meaning of either

From the Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA.

Address reprint requests to Amit Mehta MD, 8 Whittier Place, Suite 19D, Boston, MA 02114.

Copyright © 1998 by W.B. Saunders Company

0897-1889/98/1104-2004\$8.00/0

the words dictated or the text typed. NLU is a technology that can allow the computer to understand a query or statement put forth in natural human language. This means understanding not simply the words spoken, but the meaning behind them; applications are limited in dictation. NLU serves as the foundation for the voice-based help system currently under development in the Advanced Imaging Laboratory at MGH.

MEDSPEAK

The MedSpeak packaged solution consists of software and a special noise-canceling microphone. It requires a personal computer with a Pentium Pro 200 MHz processor and the Microsoft Windows NT operating system (Microsoft Inc, Redmond, WA). The package contains a 25,000 word vocabulary and averages accuracy rates of over 95% for most native North American English speakers. No special training of the software is required, although most radiologists have found that their recognition accuracy benefits from completing a short enrollment into the system where individual voice nuances are detected and stored for future analysis. The language model used in the system allows it to distinguish between words and symbols that sound alike, such as "to," "two," and "too," or the symbol colon and a patient's colon. Macros can be created to generate sections of a radiology report with a single word. There are key features of VR software (native to MedSpeak/radiology) which should be sought after within any package for deployment in a radiology department and these include the following.

1. *An open radiology information system/hospital information system (RIS/HIS).* MedSpeak includes an electronic interface to download demographic data directly to MedSpeak/Radiology and upload the finalized radiology reports to be stored directly on the RIS/HIS. To achieve this link, MedSpeak/Radiology contains a documented Hospital Language 7 (HL-7)-compliant application programming interface, with standard format and standard protocols.

2. *Prestored reports.* The individual radiologist or institution can create sets of canned text reports (a "normal chest" diagnosis, for example) and store them for later recall with a single command word or acronym. When displayed, the prestored report can be further customized or simply signed as a final report.

3. *Templates.* A radiologist can create standardized report formats and make dictation even faster for routine reports by creating customized templates and storing them for later recall. The template capability gives the radiologist the ability to create a form with "fill-in-the-blank" areas within a report and recall them with a simple "macro" command. The radiologist then simply dictates the words needed to fill in the blanks. This feature allows institutions to standardize on report formats or capture key data elements.

4. *Addendum.* MedSpeak/Radiology allows a radiologist to add additional information to a "final" or "signed" report while maintaining the integrity of the existing report. This information can be shared with other information systems, such as a radiology or hospital system, through the RIS interface agent.

5. *User-defined fields.* There are five user-defined fields that give an institution flexibility to input additional information. This information can be shared with other information systems such as a radiology or hospital system through the RIS interface agent.

6. *Barcode Interface.* MedSpeak/Radiology includes a standard barcode interface that users can modify to suit their own barcode environment.

7. *System security.* MedSpeak/Radiology's user ID and password protection at sign-on and password reentry can be used for electronic signature of reports to ensure the security of patient information.

OPERATIONS

MGH conducts more than 400,000 studies per year. The annual transcription operating cost per annum is estimated to be nearly \$850,000. The average turnaround time experienced during the previous year has been 4.1 days. With such lengthy report times, the department has suffered delayed result reporting, as well as delayed billing. When reviewing the problems associated with the conventional transcription paradigm, several operational qualities were exposed. With conventional transcription systems, an inherent bottleneck exists from "bursty" dictation traffic; this becomes a mute issue with voice dictation. Second, there is an inherent delay in final report signing due to transcription delays. These issues lead to high and variable operating costs, which are also incompatible with electronic medical record efforts.

IMPLEMENTATION

Our experience with VR technology began in November 1995 when we became an alpha test site for the IBM MedSpeak/Radiology development effort. In conjunction with IBM, by January 1996, a beta version was piloted in Teleradiology and by September 1996, we had gone live with a first product release. With an initial experience, our institution felt we were ready to embark on beginning a wider implementation, and in October 1996, the musculoskeletal division was converted to complete VR with a department-wide multiphase rollout begun by January 1997. As of August 1998, there were 35 operational workstation representing a 40% penetration within the department and producing more than 400 reports per day (Fig 1).

CHALLENGES

As users of VR technology, it is evident there remain issues regarding acceptance, as well as issues with technical improvements.

Inherent to the Technology

We will continue to see improvements in sound cards and microphones that will translate to improvements in recognition accuracy and software stability. MedSpeak is currently dependent on which manufacturer's sound card is used, but as technology improves, issues of compatibility will become less of a focus. The microphone is probably the single most critical hardware element in the total speech recognition solution. A microphone that

provides noise-cancellation technology, either active or passive, becomes mandatory, especially in high-traffic areas. Currently, our most successful recognition has been experiences with a Sennheiser microphone (Sennheiser Corp, Old Lyme, CT); however, the integrated Philips Speechmike (Philips Electronics Inc, New York, NY) microphone, including barcode, trackball, and dictation functions, is ergonomically most appealing.

Acceptance of the Technology

Obstacles that our radiology department has encountered in deployment of VR have centered around user acceptance. It is always difficult to convince users to implement a technology into their practice when utilization does not immediately benefit the end user. Initial experiences with VR always initially take longer to generate reports due to training and experience requirements. Additionally, it is difficult to align incentives as cost savings are achieved at an expense to the radiologist's time. We have found success overcoming many of these obstacles by emphasizing the support from senior management for the VR efforts, as well as eliminating many of the initial technical difficulties, including RIS integration details. We have ensured much effort has been made to manage the project and provide technical support, as well as user-friendly and convenient training sessions. Last, much of our success has been achieved by stressing the importance to senior level management of a project plan with successive removal of alternative dictation

Fig 1. VR penetration by number of studies per division.

Fig 2. Cost reduction results with VR during implementation.

systems, focused by division and providing service metrics to radiology division chiefs.

BENEFITS

Clearly, the predominant benefit of VR is decreased report turn-around times. Initial results have shown a 10-fold decrease in report times from completion of the study to finalization of the report. These reductions are bolstered by the inherent benefits through PACS. VR also serves to reduce costs with initial results, demonstrating a \$350,000 savings in the first 2 years after three phases of a five-phase implementation (Fig 2). Last, VR ultimately serves to increase report accuracy as radiologists finalize reports immediately after dictation without a transcription delay.

CONCLUSION

From our department's experience of more than 10 years of exploring a variety of VR systems, we have concluded that VR is a technology that is finally ready for prime-time use. Current VR benefits the radiology department as cited here, as well as the health care enterprise. The benefits of implementation are significant, as are the chal-

lenges. With a strong business plan and implementation proposal, VR can be rolled out in an efficient manner that is both technically and financially sound. Commitment is required from senior management to make the endeavor succeed and acceptance will be a gradual process. As radiology continues to acquire a larger percentage of the shrinking health care dollar, improved report turn-around times and decreased operating costs will become a necessity. It is our opinion that implementation of current VR technology now is essential to position radiology departments for the challenging future ahead.

REFERENCES

1. Bergeron BP: Voice recognition: An enabling technology for modern health care? Proc Am Med Info Assoc Annual Fall Symposium 1996, pp 802-806
2. Frenger P: Inexpensive speech recognition techniques. Biomed Sci Instrument 30:141-146, 1994
3. Schwartz LH, Kijewski P, Hertogen H, et al: Voice recognition in radiology reporting. AJR 169:27-29, 1997
4. Serb C: Voice recognition. Just say the word. Hosp Health Networks 72:40, 1998
5. Spacone AB: Microcomputer voice-recognition program in a hospital emergency department. J Soc Health Syst 1:111-118, 1989