Experience With Implementation of a Radiology Speech Recognition System

Jeffrey D. Houston and Frederick William Rupp

Recent advances in speech recognition technology have allowed development of computer systems for real-time radiologist-driven generation of reports. The transition to a speech recognition system is a technically complex process with many potential pitfalls that can decrease efficiency and disrupt workflow. In our recent experience with installation of such a system in an academic radiology department, factors that have worked against optimal performance have included environmental logistics, hardware incompatibilities, radiology information system interface problems, lack of suitable training, and inadequate technical support. Communication of our experience is intended to allow radiologists to anticipate complications of these systems and make informed decisions regarding the feasibility of such a system in their practices. With this information, potential buyers should be able to carefully scrutinize specifications for prospective systems and, by avoiding many of the possible pitfalls, make an easier transition to a speech recognition environment.

OVER THE PAST several years, advances in speech recognition technology have allowed development of computer systems for real-time generation of radiology reports.[1] These advances coupled with the production of increasingly powerful and less expensive computer hardware have resulted in the proliferation of physician-driven report generation software. Both positive and negative experiences with these products have been described.[2,3] Graphical user interfaces allow even those with a modicum of computer knowledge to rapidly learn how to use speech recognition software.

The desire for cost savings has encouraged installation of these systems to reduce staffing expenses by shifting the responsibility of report generation from transcriptionists to physicians. Unfortunately, the potential benefits of achieving improved clinical service and decreasing report turn-around time are at the expense of the physician's time. This has particularly serious consequences for radiologists, because dictating and editing reports already consume a large proportion of their working hours.

The following is a review of our recent experience with the installation of a proprietary speech recognition system in an academic radiology department at our Veterans Affairs hospital, which is staffed by both resident and attending radiologists. We have discovered many unforseen pitfalls that can significantly deter from the overall effectiveness of such systems and potentially disrupt a naive radiology department. The purpose of this report is to provide insight for prospective customers into the myriad of potential problems associated with these systems, so that informed decisions regarding the feasibility and mechanics of implementation may be made.

An ideal speech recognition system should require minimal time for physician training and creating individual user profiles, interact well with the existing radiology information system (RIS), and adapt to the radiologist's working environment. Daily use should introduce only minimal physician overhead beyond the actual dictation time, have a high rate of accuracy for word recognition, and make reports immediately accessible to referring health care providers.

In the appropriate setting, speech recognition systems may be able to provide some of the above attributes. In our recent experience with installation of such a system, many factors have worked against optimal performance. We describe the problems and provide examples as they relate to the following categories: environmental logistics, training, hardware, RIS interface, speech recognition software, and technical support.

ENVIRONMENTAL LOGISTICS

There is a wide range in performance of a speech recognition system, which is indirectly proportional to the amount of noise in the local environment. We have had the greatest success in small reading rooms in which there is a single system in use and a lack of extraneous personnel and ambient

From the Department of Radiology, University of New Mexico Health Sciences Center, Albuquerque, NM.

Address reprint requests to Jeffrey D. Houston, MD, Department of Radiology (UNMH 1-West), University of New Mexico Health Sciences Center, 915 Camino de Salud NE, Albuquerque, NM 87131-5336.

0897-1889/00/1303-0007$10.00/0

doi:10.1053/jdim.2000.8059

noise. In this type of environment, the system has a relatively high rate of success in accurately transcribing what is dictated. However, quite the opposite effect is seen when the system is used in a large, busy reading room. In our general reading room, where there are abundant ancillary personnel and up to 6 radiologists simultaneously dictating, moving alternator panels, talking on the telephone, conversing, and consulting with clinicians, the recognition error rate is profoundly greater, with up to multiple errors being made in every sentence.

If speech recognition is to be successful, a suitable dictating environment must be provided. Therefore, the first consideration in determining whether a speech recognition system will be feasible is if it will adapt to the anticipated physical environment or if the environment itself can be reconfigured. It is unreasonable to expect acceptable results out of a system placed in an unsatisfactory environment.

TRAINING

Adequate user training is a cornerstone for successful implementation of a speech recognition system.[4] Although a key radiologist who promotes speech recognition technology could train others, the use of dedicated training specialists is preferable. Group training may be used, although individualized training allows users to progress at their own pace and accommodates all levels of computer expertise. In addition, an ongoing training system must be created to address future system modifications and the training of newly hired radiologists and residents.

It is crucial during the user enrollment phase to train the system with the full number of sample phases, not just the minimum accepted. It is important to dictate naturally, using the same volume, cadence, and enunciation, as opposed to rushing through the voice enrollment phase. A problem we encountered with our initial enrollment was the erroneous training of the system "default queue" rather than creating personalized voice enrollment files, necessitating re-enrollment for nearly 10% of our physicians. In addition, software was not installed initially to allow the system to "learn" from the users' daily input of "problem words/spellings" and incorporate these updates into the individual profiles.

A consistent mouth-to-microphone distance is important to maintain an appropriate volume level. This is facilitated by the use of a headset-type directional microphone instead of a hand-held microphone.

Training should include the use of speech-operated commands, including the proper utilization of macros and templates, as well as any applicable programmable microphone button training. The use of macros and templates can reduce dictation and editing time, because a single trigger word can produce entire blocks of predefined text. However, difficulty abounds in naming trigger words so that the system does not confuse them with general dictation words.

A summary of user aids and shortcuts also should be posted at each dictation station, along with telephone numbers for support personnel.

HARDWARE ISSUES

During the transition from a conventional transcriptionist-based system to a speech recognition system, radiology departments may choose to gradually introduce the new system by sequentially adding workstations and users. When systems are not purchased at the same time, there is an increasing probability of hardware compatibility problems. Although computers may appear identical on the outside, their inner components may be very different. This may be inconsequential for some hardware components, but it can be a significant issue for devices such as microphones and sound cards. We discovered that in our batch of seemingly identical systems, there were sound cards produced by different manufacturers. Subtle variations in audio characteristics resulted in poorer speech recognition when used by a radiologist who had created an enrollment profile on a system with a different sound card. Steps should be taken at the time of hardware purchase to ensure that all systems have identical hardware components.

Even among systems with identical hardware, we have found that each dictation station has different microphone sensitivities. Recalibration for each radiologist before beginning a dictation session after changing workstations may be needed for optimal performance in some systems. Because this process is repetitive, cumbersome, and time consuming, it encourages use of suboptimal microphone settings and results in increased recognition errors.

Seemingly essential hardware components that can be overlooked easily during the design phase of the system are backup devices, such as magnetic

tape or optical disc drives. Reports that have been uploaded to the RIS often are safe. The unsigned reports and individual voice enrollment files, if only stored on the speech recognition system, are most susceptible to loss. Because radiologists spend many hours creating personalized enrollment files and continually train the system to understand new terms and to improve accuracy of recognition, loss of these files could be devastating. Therefore, backup mechanisms must be in place to automatically archive these files on a frequent basis.

A radiology department may not have the computing infrastructure in place to support an easy transition to a speech recognition system. The radiologists at our institution initially were pleased to see that microphones had been purchased that contained infrared bar code readers so that medical record numbers could be easily scanned off the requisitions. Enthusiasm quickly faded when it was realized that the department's fleet of aging 9-pin dot matrix printers with depleted ribbons could not produce readable bar codes, forcing manual entry of the 9-digit medical record number for each examination.

RIS INTERFACE PROBLEMS

A major selling point of different proprietary systems is the availability of a software interface from the transcription system to the existing RIS, which is the repository for transcribed reports. Although the advertised ability of a system to interface with a particular RIS is binary, there is a wide spectrum of actual performance.

A touted advantage of speech recognition systems is the immediate accessibility of reports to referring health care providers. Because of interface limitations, our reports actually have become less accessible compared with the previous telephone-based retrieval system. The unforseen factor that resulted in this reported inaccessibility is that only staff-signed reports are immediately accessible. A majority of reports are dictated by residents, resulting in preliminary unverified reports being queued in the system for signing by the attending radiologist before they are forwarded to the RIS. Thus, if an attending radiologist leaves the hospital before signing all pending preliminary reports, the unsigned reports are locked in the attending physician's report queue, and thus inaccessible to the referring clinicians. Further, physicians also are generally unable to sign reports from home or from an off-site office with speech recognition systems, whereas this is a commonly supported feature of radiology information systems. This has resulted in reports not being accessible for many days and increasing telephone calls from clinicians requesting reports on studies that already have been read but are not accessible. These reports then must be retrieved by the computer system administrator, recalled from memory by the interpreting resident, or reinterpreted, further diminishing overall departmental efficiency.

Another unforseen system idiosyncrasy that was discovered after weeks of reports were submitted was that any text on the same line as "Impression" was stripped off by the RIS, making many of our impressions unintelligible because the most important conclusion was omitted. Thus, final reports on the RIS initially should be carefully examined for subtle errors after the installation of a transcription system. Furthermore, many resident radiologists were dismayed to find that their names ironically were relocated to the "transcriptionist" field on the final report.

Hence, information regarding the exact behavior of the interface between the speech recognition system and the RIS must be sought and documented before purchase. Software modifications in the RIS may be necessary for communication between the systems. Visitation or communication with other institutions using the same speech recognition system is advantageous for verifying actual performance and discovering potential problems or limitations.

SPEECH RECOGNITION SOFTWARE ISSUES

Perhaps the most important component of the speech recognition system is the interface with which the dictating physician interacts. The system must be streamlined for efficiency and provide a compromise between unnecessary prompting and sufficient user protection. Even the most seemingly innocuous confirmatory prompts can become highly annoying and obstructive to overall efficiency after dictating several hundred reports.

The user interface should provide all of the necessary information for the interpreting radiologist in a compact and visible fashion, including the patient name, medical record number, examination description, and other examination identification information such as date and time. The names of the resident and attending radiologists should clearly

be visible to avoid confusion in a multiradiologist environment.

Functions needed by resident physicians are overlooked easily in speech recognition software. For instance, we have found that residents often dictate reports for a variety of attending radiologists in succession, but this is now complicated by the fact that our software is only able to adequately accommodate a single resident and faculty combination in one session. Furthermore, this must be defined before the report is dictated. Although the attending radiologist name seemingly can be changed midsteam, the report nevertheless will erroneously be assigned to the faculty defined at the beginning of the session. Hence, to change ownership of an already dictated report successfully, the text must be copied to the system clipboard, then the original report dictation must be aborted, the resident must then specify the new attending name, re-enter the examination identification information, and, finally, paste the text from the clipboard back onto the screen. Because this process is essentially impractical, it results in many reports being sent to the wrong faculty for signing, further delaying report availability to the referring clinicians.

If a radiologist will utilize multiple dictation stations, the system must be able to allow multiple sessions to run simultaneously. We have found that physicians who are simultaneously reading multiple imaging modalities often have to waste valuable time by retracing their steps through the department to find the workstation on which they are currently logged in, so that they can log off and start a new dictation session on another workstation. Having the option of using an automatic time-out function would be beneficial.

Perhaps the most feared event for a radiologist is the sudden loss of a long and complicated dictation during a power outage, system crash, or erroneously perceived user directive (it is astounding how many different phrases can be misinterpreted by the computer to discard a report). Automatic interval backups are an indispensable feature. To be protected from catastrophic data loss, the user must be able to step through a sequence of saved dictations in regular time intervals to be able to recover a damaged report.

Major system crashes are inevitable and result in an immediate cessation of workflow. Diagnosis of the problem and subsequent repair or replacement of proprietary hardware or software components can take several days and may require on-site service visits. Our department has been paralyzed for days while waiting for computer repairs, creating backlogs of unread studies. Therefore, a backup contingency plan must be in place. This might involve traditional transcription systems, complete spare swapable file servers and workstations, or other approaches.

TECHNICAL SUPPORT

A mandatory component of a successful speech recognition system is access to readily available qualified technical support personnel. This may require hiring a computer professional, because expertise in hardware, networking, operating systems, and software applications is needed. Having a capable system administrator allows quick resolution of problems as they arise, minimizing the impact of inevitable computer malfunctions on departmental operations.

The intention and ability of the manufacturer to provide long-term upgrades, technical support, and patches to correct programming errors are enormously important considerations regarding individual speech recognition software packages. Because the software often is the most expensive component of the system, upgrading from a discontinued product line can be a very expensive, time-consuming, and complicated endeavor.

Vendor support for our software was discontinued approximately 9 months after the introduction of the system. Without continued technical support, newly arising problems will never be solved. Now, 15 months after installation, our system is noticeably deteriorating in the absence of vendor support, which will inevitably and rapidly necessitate its replacement with a new system.

CONCLUSION

The implementation of a speech recognition system is an expensive, time-consuming, and technically demanding process. This process has the potential to become needlessly problematic without insight into the numerous and multifactorial pitfalls.

It is imperative to realize that the scope of a speech recognition system encompasses more than just the obvious hardware and software factors, but also environmental logistics, user training, and technical support issues. For the system to be

effective and nonobtrusive, all of these components must work in harmony.

From a hospital administration perspective, the experiment has been a success. Our transition to a speech recognition system has been effective in obliterating a significant backlog of transcription delays and eliminating the staffing cost of transcriptionists.

Unfortunately, the project also has been unsuccessful on many other fronts. The experience has been very frustrating for many of the radiologists who must utilize the system on a daily basis. Despite foreshortened reports, dictation time has increased significantly, thus decreasing overall radiologist productivity. Some of the attending and resident radiologists in our group, antagonized by the consistently poor rate of accuracy, now have resigned themselves to manually typing their own reports.

The difficulties we have experienced stem from a multitude of unrelated factors superimposed on a foundation of unrealistic expectations. Most of the environmental logistic problems are solvable, given sufficient resources to completely redesign existing reading rooms. Effective training and technical support come at a price. The majority of the described hardware, RIS interface, and software problems can be avoided by anticipation, careful selection of components, and meticulous attention to detail during the purchasing and installation phases. The key question is whether the inherent error rate of a speech recognition system is acceptable to the individual user, and whether significantly decreased radiologist productivity is offset adequately by the realized institutional benefits.[5]

Hopefully, communication of our experience will allow radiologists to make an informed decision on whether a speech recognition system would be suitable for their practices. Knowing the problems with our system should facilitate more careful scrutiny of the specifications and functionality of prospective systems. Radiologists who do elect to embrace this new technology will be able to make an easier transition to a speech recognition environment and avoid many of the potential complications.

REFERENCES

1. Rosenthal DI, Chew FS, Dupuy DE, et al: Computer-based speech recognition as a replacement for medical transcription. AJR 170:23-25, 1998

2. Ramaswamy MR, Chaljub G, Esch O, et al: Continuous speech recognition in MR imaging reporting: Advantages, disadvantages, and impact. AJR 174:617-622, 2000

3. Zemmel NJ, Park SM, Maurer EJ, et al: Evaluation of voicetype dictation for Windows for the radiologist. Medical Progress through Technology 21:177-180, 1996-97

4. Couris J: Road to voice recognition includes planning, training. Diagnostic Imaging PACS Supplement P35-9, Sep 1999

5. Hundt W, Stark O, Scharnbery B, et al: Speech processing in radiology. European Radiology 9:1451-1456, 1999