Speech Recognition in the Electronic Health Record (2003)

Save to myBoK

This practice brief has been updated. See the latest version here. This version is made available for historical purposes only.

Introduction

Speech recognition is not, in and of itself, the final solution in clinical documentation. Whether recognition takes place on a server in order to increase the productivity of transcriptionists or is used directly by the dictator with the goal of eliminating both the delay and expense of transcription, it should be viewed as only one component of clinical documentation. This practice brief undertakes to increase HIM professionals' understanding of how speech recognition works, the driving forces that are shaping the current and future applications of this technology, the benefits and risks associated with both "front-end" and server-based use, and to provide a glossary of terms, as well as illustrate work flow, tasks and skills, and best practices.

How It Works

Speech recognition uses mathematic probabilities of when and how often words will appear in a particular context. The acoustic model captures the acoustic properties of speech and provides the probabilities of the observed acoustic signal given a hypothesized word sequence. The language model captures the linguistic properties of the language and provides an a priori probability of word sequence, usually based on statistical concepts.

To break down sounds into written language, the speech engine takes the digitized signal from the microphone and converts it from a time-based signal into a set of frequencies. From these frequencies, the position of the vocal tract formants can be extracted and represented as a set of numbers. These numbers are then compared with a table of known formant positions for written phonemes. The formant table was developed using data captured from many hundreds of samples of native speakers of the language and creating average results. When a match is found, the corresponding phoneme is passed to the next stage of the recognition process, sentence analysis.

At this stage, the speech recognition engine analyzes the recognized words and statistically compares them to other words in the language model using a probability tree. The number of times a word appears in conjunction with other words is recorded. The analysis calculates the probability of one word following another, or appearing at the beginning or end of a sentence.[1]

User-specific training replaces the general model of speech with one that is based on the speaker's own pronunciation. The user then builds a vocabulary, which is usually built from existing documents but can be purchased for a variety of professions and specialties. Because the applications for healthcare use a different set of probabilities than those for other professions, business, or personal use, different vocabularies and language models are used to accommodate the type of language use.

Enabling computers and their applications to interact directly with human speech has many significant implications. Thanks to advances in personal computing capabilities, widespread use of speech recognition throughout the healthcare enterprise is just beginning to be taken seriously, and it is likely to give a significant boost to the goal of making 100 percent of all patient health records electronic.[2] Speech-enabling the PC is a critical step in bringing the electronic patient record closer to the physician—it allows them to continue to use speech as the fastest and most productive method of document creation. Speech is the most common means of communication between people. The promise of speech recognition is that speech will also become the most common means of communication with computers.

Driving Forces

Handwriting is becoming less and less acceptable— although it provides an immediate record, the docum- entation is frequently not as comprehensive as a dictated note, and legibility is an issue. Speech-to-text is more likely to result in a legible and comprehensive document, but there are indications of a growing shortage of transcriptionists qualified to perform the manual labor required by traditional dictation and transcription. Speech recognition used directly by the physician-dictator, in conjunction with an electronic health record (EHR) or as a background process using server-based recognition, becomes a viable option in the face of decreasing reimbursement, rising costs, growing labor shortages, and increasing demands for more complete documentation provided in a more timely manner.

When coupled with speech recognition, the EHR may provide the ideal combination of flexibility, convenience, and efficiency. It combines the best of both technologies and goes a long way in minimizing the draw-backs of each. Now in its second year of using speech recognition in conjunction with an EHR, New York-based Nassau Orthopedic Surgeons and its seven physicians estimate that practice costs are down by $100,000 (annually) and volume is up by approximately 3 percent. [3]

As with all technology, speech recognition and the hardware that supports it have improved and will continue to improve, making it a viable option to a more computer-literate generation of healthcare providers and HIM professionals.

Benefits and Risks

Speech recognition technology (SRT) has the potential to enhance clinical documentation in multiple ways. The demand for documentation with every patient care encounter is markedly on the increase. This information is needed promptly and accurately to ensure optimal patient outcomes. In addition, there are not enough experienced medical transcriptionists (MTs) to meet current and future demands.[4]

To keep up with documentation requirements, implementing SRT may be the key to making healthcare clinicians and MTs more productive participants in the documentation process and keeping pace with increased demands. There is clear interest and movement to use speech recognition in the healthcare setting. An HIMSS survey disclosed that 19 percent of IT executives are currently using speech recognition and 46 percent plan to use the technology in the next two years.[5]

In order to analyze the effect speech recognition can have in delivering increased documentation faster and more accurately, while reducing costs, one of the first steps prior to the development of a return on investment (ROI) is to assess the readiness of the medical staff in terms of their receptiveness to a transition of this magnitude. If they are proponents of full application of the technology, which means a commitment of learning to use the system and allocating resources to apply this in practical applications, ROI can be structured around an objective analysis of both the benefits and the risks.

The Benefits

In order to gain the most benefit from any technology solution, it must readily fulfill expectations of the facility's administration, medical staff, and the HIM Department. Interoperability is a key factor, and integration with the facility’s current EHR system will be essential. A facility may determine that the right approach is to implement SRT in just one department. If the results of a single-department implementation prove successful, additional departments may be scheduled to follow.

Improved Turnaround

Many facilities are experiencing transcription turnaround delays in the range of 24 to 48 hours or longer. When the information contained in the reports influences treatment decisions, the delayed dissemination of the information can hinder decision making even without prolonged turnaround. Speech recognition has the potential to improve that wait time dramatically. The physicians at Southern Hills Medical Center in Nashville, TN, are able to dictate, edit, and sign reports in one complete step within minutes from commencement to electronic signature.6 In this application, a patient's test results are often faxed to his or her physician's office before the patient has arrived home.

Reduced Costs

When MTs are used as medical text editors for a transcript generated by speech recognition on a server, reduced costs expressed in productivity gains for MTs are based on the expectation that the MT will no longer be required to manually produce the entire dictation; rather, the MT will review the voice file to the text provided and edit for missing or incorrect content, as well as format the document. Productivity gains should be measured against the generally accepted industry standard of four minutes of transcription time to each one minute of dictation, and average edit review time of two to three minutes per one minute of dictation.

Keeping in mind that documents produced by a transcriptionist will have appropriate formatting and punctuation where server-based speech recognition transcripts will not, productivity gains should be measured against these standards. Any productivity increases will be directly proportionate to factors that include quality of physician input, SRT processor recognition of input, and software application used. Essentially, use of the technology holds no guarantee that cost savings will automatically be recognized.

It is difficult to measure with any accuracy the savings to the physician for dictator-based (front-end) use, as the entire process of dictation, review, and approval of a traditionally transcribed document does not take place all in a single time block, as it does with SRT. While a physician may perceive that the process takes longer, consideration needs to be given for the fact that when the dictator finishes dictating and reviews the document, it can then be signed and distributed—time currently spent after the transcription is done. Dictating—whether in the traditional way or to SRT—is not only less time-consuming than handwriting, but typewritten records are legible and usually more detailed and complete. SRT has the capability of enhancing physician productivity, leaving more time for direct patient care.

Error Reduction

Editing text, whether done by physician or MT editor, reduces content errors in patient reports, provided it is done meticulously prior to signing. In current transcription practices, many transcribed reports are not reviewed closely before a signature is applied by the physician. Standards for ensuring accuracy with all documents produced using speech recognition call for third-party editing.[7]

Improved, Timely Medical Decision Making

The medical decision-making process is optimized by information. The use of speech recognition can reduce the amount of time it takes for information to be made available to other healthcare providers. What may have taken hours in a traditional dictation setting can be accomplished in minutes (user-based SRT) or in a shorter amount of time (server-based SRT). For example, a single-step process such as a radiology or an emergency department[8] illustrates the time savings. In serious trauma cases and critical care cases, prompt and accurate medical treatment determinations can not only save lives, they may substantially improve patient outcomes and reduce patient care days as well.

MT Transition to Editing

Initially MTs may naturally be apprehensive about the use of SRT. They may even be resentful or hostile. Transcriptionists are poised to grow into evolving clinical data specialists, data quality managers, and decision support specialists, as envisioned by AHIMA's Vision 2006 initiative.[9] Moving to clinical data editors (or whatever future title may emerge) is a process that recognizes and values MTs for their expertise and skill to interpret the subtleties of language. It will be crucial to bring about an understanding that the benefit for both MTs and the healthcare facility is to enable them to keep pace with increased documentation requirements.

The Risks

While SRT sounds like the panacea to all medical transcription backlog woes, including rising costs and the reported shortage of qualified MTs, there are some technology shortcomings and risks to consider when making the decision to use speech recognition.

The speech application market originally targeted client-side dictation; however, this market has been slow to get off the ground because usability issues prevented the technology from offering improved productivity. In a study of participants who were native English speakers with good typing skills, the fastest users spoke an average of 107 uncorrected words per minute, which resulted in approximately 25 corrected words per minute. The “keyboard-mouse” group completed almost three times more words per minute than did the “voice-only” group. Participants observed that they were usually aware when a typing error occurred, but were much less confident about being aware when a speech recognition error occurred.

The study concluded that users must either constantly glance at the display for errors or rely heavily on proofreading after the speaking has ended.[10] (Note: Many of these studies were authored in the period from 1999 to 2001; more current studies could not be located, but the technology has advanced and will continue to advance, so studies become quickly outdated. This particular study, however, was also cited in a 2002 study.)

Time Is Money

The cost in time to the physician-dictator in using front-end speech recognition is most likely based on more than perception. Hospitals and group practices looking for ways to get physicians to dictate, format, correct, and self-edit their documentation will have to show clear value.[11] If there is no clear value, such as the immediate availability to the record, it is questionable whether physicians will embrace a process that takes time from patient care.

Editing Costs

Unless the recognition accuracy is very high and the software package has been enhanced to speed the process, the amount of time it takes to edit and form at a document transcribed by server-based SRT could exceed the time it takes to transcribe manually. Current technology will not generate an acceptable level of accuracy for all users, which will require either continued manual transcription or combined use with a system that reduces the amount of free- text dictation (templates, EHR, etc.).

Costs

SRT can be a costly investment. Before decisions are made regarding such capital expenditures, a facility will need to look at many options, consider varying technology solutions, and explore future upgrades to the technology as well as maintenance costs. Additionally, optional enhancements such as networking, hand-held device usage, and the system's ability to integrate with the hospital central system to provide upgraded tools such as e-sign and auto-fax need to be evaluated.

Technology Mismatch

A technology that does not align with an organization's needs could be catastrophic. Having the support of administration, and especially the IT Department, in adopting SRT will be a determining factor in the potential success of the project. When planning to implement SRT in any form, identify what the users' expectations are in terms of input, time of usage, and willingness to be trained, and obtain a commitment from all stakeholders to use the system until the output quality has reached expected levels. Selecting a technology that is scalable to the expectations and widespread usage envisioned in the initial ROI will be important in selecting and deploying the technology. Investing in a system that does not become fully used may be worse than making the decision not to apply the technology at all.

Edits and Content

It is recommended that edit review of every document be implemented to ensure accuracy. The time it takes to edit the reports and ensure that all information has been captured correctly negatively affects the advantage of having a document ready for distribution when the dictator is finished speaking. Additionally, prompts are not available with server-based SRT or some applications; required section heading content and formatting may be inadvertently omitted, so critical data capture may be missing from reports, which would require additional dictation or addenda (if the report is already signed). SRT usage does not overcome disorganized dictation, poor grammar, or missing or overused punctuation.[12]

Who will edit reports? Each facility will need to make this determination based on the applications being integrated with the SRT. In some cases, physicians are willing to take on this task to fully own and manage the process from beginning to end and to have the ability to disseminate the document immediately. However, because the physician is the most expensive individual in the hospital, this decision requires careful consideration. Medical staff has to be willing to take on this responsibility, especially taking into consideration that editing time takes physicians away from providing primary patient care. Having an MT text editor affects turnaround time to the chart. Every facility considering SRT implementation needs to fully review the options and their implications before investing in hardware, software, and training.

Using Speech Recognition

End-user or Front-end Speech Recognition

"Front-end" speech recognition is the term generally used to describe a process where the dictator (end user) speaks into a microphone or headset attached to a PC. The recognized words are displayed as they are recognized, and the dictator is expected to correct misrecognitions.

The advantage is that the dictator is in control of the entire process—the document is dictated, corrected, and authenticated all in one sitting. When dictation is done, the document is ready for distribution. Front-end speech recognition is also the most effective use of SRT with an EHR, enabling the dictator to respond to prompts from the EHR for more complete and accurate documentation.

End-user speech recognition may affect a dictator's billable activities, however. Training the speech recognition engine is a time-consuming process that takes time away from patient care. Furthermore, dictators are distracted when they read the on-screen speech translation because the system revises the interpretation as it goes—watching the changes is distracting and slows dictation.[13] The dictator is also performing the duties of an editor, as any requirement to send the document to be edited by a third party negates the advantage of being able to distribute the document as soon as dictation is complete.

Even with 98 percent accuracy, one of every 50 words is misrecognized, requiring the dictator to make a correction. Failure to make corrections can degrade the overall accuracy of the dictator's language model, as the program "learns" and uncorrected misrecognitions are entered into the language model as being correct. In the amount of time it takes to make the corrections, a clinician can see three additional patients.[14]

Server-based (Back-end) Speech Recognition

Server-based speech recognition takes place after the dictator has created audio input in much the same way as usual, and the process then takes place at the server level, or on the "back end." All speech recognition programs currently on the market have the capability of transcribing a recording for an enrolled user. The end user could, in fact, record audio and use the transcribing function of the application, then edit the final document. In most cases, server-based speech application refers to a speech recognition engine processing the audio to text, sending the draft text and a synchronized speech file to an editor for correction and formatting, and then inserting the document to continue the work flow.

The advantage of server-based speech recognition is that it does not affect the end user in terms of dictation habits or time—the dictator continues to dictate as always. It also has the potential to make editors more productive, requiring fewer people to generate more documents. The time commitment to training the speech recognition engine is taken from the dictating physician and placed on individuals who are not under pressure to provide direct patient care. The captured audio file can be used to train and retrain the SRT engine for better recognition in a shorter time frame.

While server-based SRT seems to be most attractive to physicians in terms of clinical documentation, unfortunately it has some major disadvantages to others in the documentation chain. The first is that, without extremely good recognition accuracy and appropriate editing tools, documents produced may require more time to edit from the synchronized audio file than if they were just transcribed.

Speech recognition engines have limited capability to understand complex commands on the server. "Period," "new paragraph," "new line," and "comma" would all be recognized, but template fields would not. A document with no punctuation, no formatting, and 90 percent to 95 percent accuracy requires extensive editing. Studies done at Mayo Clinic in Rochester, MN, have concluded that there was no productivity gain with server-based speech recognition.[15]

The end user also has no incentive to change any dictation habits because she or he does not see the end result or have to fix it. Instructions to "go back to the history and take out the part where I said <...> and insert <...> instead" will be transcribed verbatim. Commercial speech recognition engines have programming to eliminate "um" and "uh" from the text, but users report that valid words are also dropped when this feature is activated.

Essential components that make an EHR attractive are also lost in server-based SRT. If documentation improvement is the goal, server-based speech recognition does not do anything to move a dictator toward that goal. (See also Appendix A.)

Templates and Macros

Dictating "free text" lends itself to more errors. Taking advantage of available technology, end users can improve their recognition accuracy and effectively reduce dictation time.

A template is a standardized document outline that includes any number of elements. Some companies sell templates, value-added resellers (VAR) will develop templates for a user, or a user can write his or her own. In speech recognition, a template includes fields that enable a user to skip from one field to the next using speech commands.

Macros are a series of keystrokes and/or commands that are executed on command. Speech programs are especially suited to use with macros to generate large amounts of text using only a few commands that are easily recognized. Radiology has adapted readily to speech technology because of the limited amount of terminology, but also because of the large number of "normal" results, which can be programmed as macros.

The intelligent use of templates and macros facilitates end-user acceptance of speech recognition as a device to spare more time for patient care while creating more complete documentation, faster. Use with an EHR that has been carefully selected with speech activation in mind accomplishes the same goal, allowing the clinician to document the record completely, accurately, and in a timely manner while not detracting from the primary purpose of patient care.

Equipment

Audio Input

There are specific audio input requirements for successful speech recognition. The best audio input takes place on the same sound card the recognition engine will use for transcription, but in most clinical settings this is not going to be the case. High-quality handheld microphones, headsets with attached boom, and array microphones for hands-free or headset-free dictation provide the best audio input for front-end speech recognition. Hand-held digital recorders, PDAs equipped with dictation modules, and tablet PCs can generate acceptable digital audio files for speech recognition. All devices should be noise-canceling devices or the recognition accuracy will be degraded.

Telephones do not have sufficient quality microphones, and phone lines are subject to interference, resulting in a degraded audio file and recognition accuracy. Attempting server-based speech recognition using dictation phoned into a digital dictation system would degrade the recognition accuracy.

Processor, Memory, and Sound Card

Speech recognition is a CPU-intensive process, whether it takes place on the server or on a dictator's PC. All processors and sound cards are not alike, and most speech recognition companies have specific requirements for what works best with their product. If you are considering using speech recognition, do not purchase computer hardware until you have consulted with a VAR with experience in speech recognition or the software company (see "Critical Elements, below").

Definition of Accuracy

In theory, speech recognition should be held to the same standards of accuracy as medical transcription. In practice, clinicians are willing to accept certain errors in exchange for the benefits speech recognition delivers. Each facility needs to define acceptable standards of accuracy for all documentation, whether it is handwritten, checked off a form, dictated as free text, dictated for processing by speech recognition (front end or server), or entered into an EHR by keyboard or speech commands.

Critical Elements

The critical success factors outlined below contributed to providing the following benefits to facilities deploying speech recognition:

- Improved the level of success realized
- Minimized the risks associated with such a project
- Provided a smoother transition from the legacy system

Critical Success Factors

1. **Define** measurable objectives prior to implementation.
2. **Establish a target ROI**, including time frame for achievement.
3. **Secure** executive sponsorship.

4. Actively **involve** users from all levels throughout the project.
5. **Designate** both a technical and functional system administrator.
6. **Identify** key benefits for end users.
7. **Align** the MT's compensation with the new technology.
8. **Develop** an operational plan in advance.
9. **Provide** key stakeholder updates regularly during the project.
10. **Establish** benchmarks prior to deployment for postdeployment analysis and comparison.

(See also Appendix B, “Best Practices for Using Speech Recognition”.)

Process Owners

Physicians

Physicians want to save time; institutions want to save money.[16] Physicians do not want to wait for a slow computer or deal with poorly maintained and aging hardware. They do not want to wait for a workstation any more than they want to hear a busy signal on the phone. Training must be focused, relevant, and efficient to keep and hold their interest. Otherwise, they will resist spending the time to learn a speech recognition program and edit to their own dictation.

Transcriptionists-Editors

Many transcriptionists view editing as boring and tedious and have no desire to edit. Because server-based SRT very much epitomizes the phrase "garbage in, garbage out," there may well be no productivity improvement for the transcriptionist-editor, and there will most certainly be a productivity loss during the training period. Someone who is on production pay will not be tolerant of anything that cuts into productivity for a prolonged period of time. Again, training is a key element to success. Anyone who sees the EHR and/or speech recognition as a threat to his or her job will also need the necessary training to be successful and to accept implementation of the technology.

HIM Department

The HIM department has the responsibility of developing and securing approval for the many policies and procedures surrounding medical transcription content, process, and requirements. Whether in the traditional setting or with the implementation of SRT, these policies must be clearly written and in current practice.

The HIM department routines and processes may remain very much the same with the implementation of SRT. The impact to the department related to tasks and processes is anticipated to be minimal. Traditional tasks such as charting reports and deficiency analysis for both missing dictation or unsigned reports remain the same with either technology and will continue to need to be done. HIM staff will likely be involved with training physicians and other clinicians on the use of the new technology as it relates to record completion, and effective training will take time.

However, other time factors can lead to overall department efficiencies directly related to productivity gains with SRT.[6] One potential positive effect of SRT is that there may be fewer unsigned reports if the dictator originates and completes the dictation at one time. A possible outcome of the productivity gains with SRT may be fewer reports missing from the chart at the time coding is done, leading to reduced bill-hold days, fewer number of medical records in an incomplete status, and reduced number of days records are incomplete.

IT Department

Speech recognition is a technology-heavy application and requires excellent technical skills to implement the tech-nology and support the hardware and configuration re-quirements. IT will be a key stakeholder in the use of SRT.

Evaluating ROI

Several factors affect the ability to clearly evaluate ROI:

- The EHR is still an emerging technology.
- The conditions for successful speech-to-text use for SRT have only recently made its widespread use possible, and therefore it is also considered an emerging technology.
- The number of EHR vendors, variety of applications, and disparate features offered by them
- Relatively low number of installations
- The different situations in which these technologies are used
- The human factors affecting implementation, training, and retention

(See also Appendix C, "Tasks and Skills List."

References

1. "Guide to Speech Recognition." *PC Magazine*, December 1998 Special Supplement.
2. Essex, David. "Taking Dictation into the 21st Century." Healthcare Informatics 16, no.7 (July 1999): 61–65.
3. Gainer, Cassie. "Voice Recognition: With Improved Technology, Efficiencies Are Clear." *Physicians Practice: The Business Journal for Physicans* 13, no. 2 (2203):82-84.
4. US Department of Labor, Bureau of Labor Statistics. Occupational Outlook Handbook.
5. Healthcare Information and Management Systems Society Leadership Survey, 2002.
6. Case study by Dictaphone Corp. Southern Hills Medical Center, located in Nashville, TN, is part of HCA.
7. American Society for Testing and Materials. ASTM E31.22 Standard Guide to Speech Recognition Products in Health Care [draft].
8. Zick, R., and J. Olsen. "Voice Recognition Software Versus a Traditional Transcription Service for Physi- cian Charting in the ED." *American Journal of Emergency Medicine* 19, no. 4 (2001).
9. American Health Information Management Association. *Evolving HIM Careers: Seven Roles for the Future*. Chicago, 1999.
10. Karat, C. M. et al. "Patterns of Entry and Correction in Large Vocabulary Continuous Speech Recognition Systems." In: *Proceedings of the Conference on Human Factors in Computing Systems, Pittsburgh, PA*. New York: ACM, 1999: 568–75.
11. Goedart, J. "Speech Recognition Technology Gives Voice to Clinical Data." *Health Data Management* 10, no. 12 (December 2002): 30–32, 34, 36.
12. ASTM.
13. Zafar, A., J. M. Overhage, and C. McDonald. "Continuous Speech Recognition for Clinicians." *Journal of the American Medical Informatics Association* 6, no. 3 (1999): 195–204.
14. Terry, Ken. "Instant Patient Records and All You Have to Do Is Talk." *Medical Economics* 76, no.19 (October 11, 1999): 101–102, 107–108, 111–112.
15. Derynck, A., P. Olevson, and B. Owen. "The Jour ney Continues: Server-based Speech Recognition." Proceedings of AHIMA's National Convention, 2002.
16. Hier, D. "Physician Buy-in for an EHR." *Healthcare Informatics* 19, no.10 (October 2002): 37–40.
17. Department of Computer Science at the University of Massachusetts at Boston.

Prepared by

This practice brief was developed by the following AHIMA e-HIM workgroup:

John Beats
Kathy Brouch, RHIA, CCS (staff)
Linda Bugdanowitz, RHIA, CHP
Mary Johnson, RHIT, CCS-P
Nancy Korn-Smith, RHIT
Susan Lucci, RHIT, CMT
Pamela Oachs, MA, RHIA
Sharon Rhodes, RHIT, CMT, CPC
Harry Rhodes, RHIA, CHP (staff)

Greg Schnitzer
David Sweet, MLS (staff)
Christine Taylor
Claudia Tessier
Joe Weber
Michelle Wieczorek, RN, RHIT
Julianne Weight

The Speech Recognition in the EHR work group was supported by a grant to the Foundation of Research and Education

Source: AHIMA e-HIM Work Group on Speech Recognition in the EHR. (October 2003).