2 0

05

From the February 2005 Issue

Is Speech Recognition the Holy Grail?

Speech recognition technology has been "two years away for the last 10 years." Here is a look at the issues and challenges that remain to be overcome.

By Nick van Terheyden, M.D.

Nick van Terheyden, M.D., the chief medical officer for Philips Speech Recognition Systems, is a physician and informaticist, and the author of numerous articles on healthcare IT, handwriting recognition and speech recognition. Contact him at nick@vanterheyden.com.

Speech recognition technology has been lauded as the best thing to happen to healthcare technology since the advent of the computer, but is it really the Holy Grail?

Speech recognition has the potential to overcome one of the most significant barriers to implementing a fully computerized medical record: direct capture of physician notes. Industry estimates from physicians and chief information officers at hospitals suggest that 50 percent of physicians will utilize speech recognition within five years. Coupled with this is the growing demand for medical transcriptionists, which is projected to grow faster than the average of all occupations through 2010.

Demand is fueled by two factors: an increase in the elderly population who receive proportionately greater numbers of medical tests, treatments and procedures that require documentation; and the continued need for electronic documentation that can easily be shared among providers, third-party payers, regulators and consumers. The demand for documentation with every patient care encounter also is markedly on the increase. Patient information is needed promptly and accurately to ensure optimal patient outcomes.

Caveat Emptor

Speech recognition technology is supposed to afford the identification of the spoken word by the application of sophisticated technology, but let the buyer beware here. If one believes he has only to speak and his words will appear as spoken on the computer screen, he will be disappointed. Technical problems exist that impact on the accuracy of translating voice to data. Some of the key technical problems in speech recognition include:

- interspeaker differences;
- resolution of ambiguity;
- the need to separate speech from background noise;
- punctuation and grammar rules needed in final documents.

Most current speech recognition engines are literal translation engines. Even if they are 100 percent accurate, all they produce is an exact reproduction of the spoken word. Most physician dictations are not structured to match the final report in verbal form. The literal interpretation of the dictation produces the content that makes up a report, but in nonlinear order and often including additional phrases and comments that must be edited out by transcriptionists. This negatively impacts the potential efficiency gains.

The speech engine will compare the original text against the edited version and "learn" the corrections, thus improving the overall recognition rate.

Accent and pronunciation variations of English also are impediments to reliable recognition. For example, the word "respiratory" is pronounced very differently depending on the origin of the speaker. Interpretation of words and phrases such as "to," "too" and "two" are virtually indistinguishable to this tool, yet the meaning of these three words can impact the outcome of the desired text.

The need for grammar and punctuation compounds the problem. Written language may need punctuation—commas, periods and quotation marks—according to strict rules that are not obvious in speech and are difficult to infer. Some available systems today use the statistical language model. While this is cumbersome, since it requires a huge database to back it up, it is the most robust, single tool for accuracy in converting voice to data. It does require extensive processing to calculate conditional probabilities. However, with this in place, the system is able to rule out "10 mouse" as a possibility and instead pick "10 mice."

Server-based Technology

There are voice recognition systems in use today that appear, on the surface, to provide rapid speech conversion to data. One example is server-based speech recognition that offers a win-win solution for the provider and the organization's need for improved turnaround time.

Server-based speech recognition technology processes the speech through a speech engine and converts the speech to text. The advantage of server-based speech recognition is that it does not impact the current physician process in terms of dictation habits or time. Physicians who want to avoid the tedium of front-end correction with dictation are offered the option of reviewing the speech-processed text and either performing the self-editing and signing the document or sending the document to transcription for editing and document completion.

Speech recognition has the potential to overcome one of the most significant barriers to implementing a fully computerized medical record: direct capture of physician notes.

Speech accuracy is enhanced by the system's ability to reprocess the audio file multiple times, necessary for converting the audio to text. This allows for proofing and final formatting done by the transcriptionist. The good news here is that the speech engine will compare the original text against the edited version and "learn" the corrections, thus improving the overall recognition rate. This feature enables those physicians whose voice profiles have become sufficiently accurate to avoid the need to send the speech-processed text to transcription. Routing rules are in place to make sure that the file is available to both the physician and the transcriptionist so that the document can be completed in the most efficient manner.

Cost Factors

There are three major costs of creating documentation: the cost of physician time, transcription costs and the cost of the time that the document is not available for patient care. The time and expense of any other people involved such as scribes, transcriptionists, editors, and those who print the reports, pull charts and refile these charts should also be included.

Cost to physician and/or the organization include the lost-opportunity cost spent documenting, answering phone calls when the document is not available and waiting for documentation. This time could be better spent seeing and treating more patients or performing more procedures.

	Dictation Time	Sign Time	Physician Total Time	Average Report Length	Yearly Number of Operative Reports	Transcription Time	Turnaround Time
Traditional Transcription	2 minutes	2 minutes	4 minutes	20 lines	450	8 minutes	24 hours
Speech Rec. With MT Editing*	2 minutes	2 minutes	4 minutes	20 lines	585 reports	5.6 minutes	17 hours
Savings/Productivity Increase	0	0	0	0	135 more operative reports	2.4 minutes	7 hours earlier

* Data presumes an estimated 30 percent increase in medical transcriptionist's productivity.

Since recent trends have moved speech recognition to server-based voice recognition, this does not impact the current physician process in terms of dictation habits or time. Physicians can continue seeing patients without diverting time to clean up dictations. For those physicians with the desire and with high levels of accuracy, this offers the option of real-time speech recognition and correction or reviewing the speech-processed text and either performing the self-editing and signing the document or sending the document to transcription for editing and document completion.

More importantly, delays in receiving a transcription not only have an impact on workflow and chart-searching but also may have an impact on quality of care delivered, which can directly affect patient outcomes, resulting in increased costs associated with risk and liabilities. For a referral practice, delays can affect the ability to get timely authorization for subsequent treatment.

Consider the time spent obtaining an inpatient cardiology consultation. By the time the consultation is dictated, typed and returned to the inpatient ward, at least one day has elapsed, which equals one more day of hospitalization before the patient's physician has the information to make further treatment decisions.

Impact Upon Medical Transcriptionists

All of this has affected the medical transcription industry. Transcription is considered a growth industry by the Medical Transcription Industry Alliance. The Alliance estimates that spending by healthcare facilities for medical transcription services will range from $10 billion to $24 billion annually, with an estimated 30 percent growth rate.

Turnaround time is a major hidden cost and an inefficiency in transcription. Consider this example: A practice that sees 50 patients a day and has a three-day turnaround time for completion of documents equates to 150 charts being held out for letters. The alternative involves filing 150 charts, pulling them when the transcriptionist returns and then refiling them.

Implementing systems is an expensive initiative in an already financially constrained healthcare industry. Purchasing costs for a voice recognition system include equipment, software, implementation and training for both the physician and medical transcriptionists (MTs). To be truly accurate, a return on investment analysis must allow for including assessments of time spent by both the MTs and the physicians. If the physicians perceive that the project is merely a cost savings effort on the part of the facility, there may be a high level of resistance on their part. If MTs suspect that speech recognition is a threat to their job security, there is also a possibility of resistance to adopt the change in skill sets from listening and transcribing to listening, reading and editing.

The transcription industry estimates four minutes of transcription time to one minute of dictation. This industry estimates a 50 percent increase in productivity for transcriptionists based on a physician recognition rate of 95 percent accuracy.

The earlier chart reflects only a change in the productivity and accuracy of the medical transcription and does not show a positive impact on the physician's time.

Reality Check

While some factual background points to the positive impact of speech recognition technology upon the healthcare industry, there must be further improvements if this tool is to succeed in being an adjunct to successful electronic documentation. Areas that need improvement include:

Noise correction software. One approach to separate the user's speech from background noise is to build up databases of what background noise sounds like, which means it can be identified in the same way that speech is, and then it can be eliminated.

Enhancing dictation devices. Work is under way to improve the microphones used in speech recognition systems. Microsoft believes that using more than one microphone will help a speech recognition system distinguish between noise and actual speech.

Detection of vocal chord movement. Another theory being explored by Laurence Livermore labs is to use radar to detect movements of vocal chords. These signals can be added to the speech engine and processed in a similar fashion as sound waves and discriminate between noise and voice.

Natural language understanding. The merger between natural language processing (NLP) and continuous speech recognition (CSR) represents the most exciting potential advance in this area to add real value to the clinical process and extract clinical content and data from conversational clinical dictation.

A Hopeful Future

Speech recognition is good technology, but it is neither a panacea nor the Holy Grail. Speech recognition has been two years away for the last 10 years, but we may be approaching the Grail—finally.

Developments over the last several years have incrementally improved speech recognition systems to the point that some have intelligent speech interpretation—extracting the meaning, not just the literal translation of words—and producing high-quality documents with minimal human intervention. Further integration and embedding speech recognition with mainstream EMR solutions will allow for expedited capture of documentation as part of the clinical care process, offering clinicians a choice of methods to document creation.

The last significant development in speech recognition technology was the recognition of continuous speech. The next big leap in this technology will be the merger of NLP and CSR to create natural language understanding. This development will take the technology to the next level and will offer a realistic opportunity to make speech recognition the de facto method of data capture for the medical community. The question is, When?