Medical Dictation: A New Generation

by Gabriel F. Groner, PhD

As the term implies, automatic speech recognition is a computer's recognition of speech, automatically converting it into readable text and interpretable commands. To fully understand automatic speech recognition, one must first be aware of the distinction between speech recognition and voice recognition, terms that are often used interchangeably. Technically speaking, voice recognition refers to recognition of voice, or who is speaking. Its purpose is to identify the speaker or to verify for security purposes that the speaker is who he claims to be. Speech recognition is the recognition of content, or what is spoken. It conveys commands and information. And after several decades of development, speech recognition technology has become a practical tool for creating electronic patient records.

Applying It Today

When it comes to giving commands and entering data at workstations, speech recognition technology has been used in a practical manner for several years. For example, users who combine this technology with word processing applications can state commands such as "open the document called schedule" or "underline the last sentence" as they work on a document. Speech recognition—when used with appropriate application software—has some advantages over keyboard and mouse technology: commands can be intuitive, more flexible, and require fewer steps. Furthermore, speech recognition does not require visual data entry fields on a screen. Rather, the person entering the patient's healthcare data can create a form, then select certain characteristics about the patient using verbal statements. For instance, a healthcare professional entering information on a patient who smokes could state "go to smoking," then click the item once it appears on the screen, thus pulling the information into the patient's form. Or one could simply say "open patient history for 553873456," followed by "smokes two packs a day" to transmit the information. Currently, most speech recognition technology for commands and data entry works with any person—without adjusting for individual voices. Most Americans have already encountered this technology on the telephone, whether it is selecting from spoken menus, checking bank balances, or trading stock. One can see how this would easily translate into entering healthcare data over the telephone.

How It Works

The widespread availability of inexpensive, high-speed personal computers with large memories combined with improvements in speech recognition technology now make it practical to automatically produce draft transcriptions from continuous dictation. The speech recognition technology for these short commands and phrases typically recognizes hundreds of words in predictable sequences. However, the technology for medical record dictation must recognize tens of thousands of words spoken in any reasonable order. It also must recognize commands. Because of its complexity, the latter form of technology imposes practical limitations. In particular, it must adjust to individual voices, requires high-quality sound, and works only for predefined vocabularies.

When transcribing dictation, speech recognition software recognizes sequences of phonemes (the basic speech sounds that distinguish meaning) and divides predefined phoneme sequences into words. To help users attain high recognition accuracy, speech recognition systems include summary information about how large numbers of individuals pronounce phonemes. Furthermore, individual users must adjust this information, through a "training" or "enrollment" process, to accommodate his or her pronunciation. Enrollment usually requires the user to read aloud to the system. Training automatically continues to improve recognition accuracy as the product is used. The time spent enrolling is worth the effort, since bypassing it typically leads to disappointing results for the user.

Automatic recognition of tens of thousands of words—many of which are distinguished by high-frequency sounds—requires good microphones and good audio computer cards or other audio recording equipment. Tape recorders and other recording equipment that physicians have used in the past are generally inadequate. Currently, automatic dictation via ordinary telephone

lines is unreliable. However, digital recorders that meet speech recognition requirements are becoming practical and available. Recordings can be transferred manually or over telephone lines to computers that recognize them. Then they are transcribed automatically in batch mode—as needed or as processing permits.

Speech recognition products usually recognize general English, business English, or general medical vocabularies—each consisting of approximately 25,000 words. However, these limited vocabularies are inadequate for medical professionals who dictate proper names and practice-specific vocabulary words. To make up for this, speech recognition systems offer methods to expand their vocabularies. The most basic method is for the user to dictate for many hours, but this frequently results in poor recognition accuracy until the software "learns" the new vocabulary. Users also can build vocabularies by collecting and processing a great deal of representative, computer-readable text. Alternatively, users can purchase specialized vocabularies, which they can further customize by adding proper names and any other practice-specific words.

Text formatting or "style" is another aspect of speech recognition that users must consider. Manufacturers generally follow guidelines established by professional transcription organizations so that text is presented in a clear, concise, and consistent manner. Just as with professional transcriptionists, creating speech recognition vocabularies that conform to style guidelines will result in text that requires minimal correction. Obviously, vocabularies that do not conform to these style guidelines will lead to text that must be corrected both for style and recognition errors, thereby increasing transcription costs.

For Best Results...

To function at top capacity, speech recognition technology and vocabularies must be used in conjunction with a software application. At a minimum, that application should be a word processing program such as Microsoft Word or Corel WordPerfect. But applications that provide customized templates will save a great deal of time, as clinicians need only dictate patient-specific information and abnormal observations. Applications that enable patient records to be manipulated by voice let clinicians verbally locate a specific section for a particular patient, speak the data values, and dictate narrative findings and orders.

Used correctly, speech recognition products can save time and money. Experiments comparing the time required to manually transcribe dictation versus the time involved in correcting draft transcriptions produced with automatic speech recognition show that the latter saves time if its recognition accuracy is sufficiently high. When approximately 25 percent of words are recognized incorrectly, it takes as long to correct recognized text as it does to transcribe the same dictated text. However, it only takes half the time to correct a document as it does to transcribe it when the recognition error rate drops to 5 percent. Furthermore, speech recognition has greater difficulty recognizing short words—articles, pronouns, and prepositions—whereas medical transcriptionists have most difficulty with medical terms such as drug and procedure names. Therefore, as the use of speech recognition technology increases, the role of the transcriptionist will change to that of an editor, rather than a typist. The skills of the medical transcriptionist will encompass editing the document and verifying that the information is correct.

Into the Future

Speech recognition technology and its applications have only recently become practical and will continue to require less adjustment and customization for individuals. Digitized speech will be transmitted over the Internet so clinicians can dictate from anywhere, just as they do now. Applications will be easier and more efficient to use as they match clinicians' needs and work styles better. They will include report formats, normal responses, and database references to minimize dictation. Finally, use of speech recognition technology will enable healthcare professionals to include more narrative text in electronic patient records, thereby increasing the amount of accessible, useful clinical and administrative information.

Gabriel F. Groner *is chief technology officer at KorTeam International Inc. in Sunnyvale, CA.*

Article Citation:
Groner, Gabriel F. "Medical Dictation: A New Generation." *Journal of AHIMA* 70, no. 6 (1999): 20-21.