Evaluation of the Accuracy of a Continuous Speech Recognition Software System in Radiology

Kalpana M. Kanal, Nicholas J. Hangiandreou, Anne-Marie G. Sykes, Heidi E. Eklund, Philip A. Araoz, Jorge A. Leon, and Bradley J. Erickson

RADIOLOGY REPORTS in most medical settings are generally dictated by the radiologists and then transcribed by a human transcriptionist, resulting in a text report. The radiologist then finalizes the transcribed report after reviewing it and assuring the accuracy of the text. Time delays between the various stages of this process usually mean that the final reports are available only after several hours or more have passed following interpretation of the examination.

The emergence of automatic speech recognition software has suggested that all reading rooms operate in the direct dictation mode without involving the human transcriptionist. When used in conjunction with electronic systems for managing the text information (radiology information system [RIS]) and image information (picture archiving and communication system [PACS]), speech recognition software may allow all finalized radiology examinations to be delivered to clinicians within minutes of interpretation by the radiologist.

Early speech recognition software products required the user to speak in a discontinuous manner, so that each individual word could be identified and transcribed.[1-7] Overall accuracy, as determined in one study of a discrete speech recognition system, was reported to be 97.6%.[7] The requirement for discontinuous speech made these products impractical for routine use in a high-volume radiology reading room. Newer products allow the user to speak in a more natural, continuous manner.[8] Our aims in the current work include measurement of the accuracy of one continuous speech recognition product, investigation of the impact on accuracy of the gender of the speaker and status of the speaker as a native or non-native English speaker, and evaluation of the potential for routine clinical use of the system for radiology report transcription.

METHODS

IBM MedSpeak/Radiology software, version 1.1 (IBM Corporate Offices, Armonk, NY) was evaluated. This software allows continuous speech to be transcribed to text as it is spoken. Six speakers, three males and three females, familiar with medical and radiological terminology participated in the study. Two of the speakers were non-native English speakers. Each speaker performed the minimum enrollment (training) procedure, and dictated a set of 12 preselected reports. The reports included neurologic and body imaging examinations performed with six different imaging modalities.

Once the original and dictated reports were compared, each discrepancy was classified as one of four different error types. Class 0 errors involved no change in meaning with respect to the original report text, and the transcribed text was grammatically correct. Class 1 errors also involved no change in meaning, but the transcribed text was grammatically incorrect. Class 2 errors were those in which the meaning of the transcribed report text was different than that of the original report text, but the error was judged to be obvious. Class 3 errors also involved a change in meaning as compared with the original report text, but the error was judged not to be obvious. In general, a single error could consist of either a single word, or a multiword phrase.

Once the errors were classified, error rates for three categories of error were computed for each dictated report by dividing the total number of errors qualifying for each category by the total number of words in the report. "Overall errors" included all four error classes (class 0, 1, 2, and 3). "Significant errors" included only class 2 and class 3 errors. "Subtle significant errors" included only class 3 errors. The dependence of the error rates on imaging modality, native English speaker status, and gender were evaluated by performing t tests using a 95% confidence level.

RESULTS

No statistically significant differences between the overall error rates for the different types of reports were observed and thus the error rates for each speaker and error class pooled across modality were computed. Pooling across the entire group of six speakers, the error rates of overall errors, significant errors, and subtle significant errors were found to be 10.3% ± 3.3%, 7.8% ± 3.4%, and 1.2% ± 1.6%, respectively.

The native English speaker error rates are all lower than the corresponding error rates for non-native English speakers, and these differences were found to be statistically significant for the overall

From the Department of Diagnostic Radiology, Mayo Clinic, Rochester, MN; and Department of Radiology, University of Minnesota, Minneapolis, MN.

Address reprint requests to Kalpana M. Kanal, PhD, Deaprtment of Radiology, University of Minnesota, 420 Delaware St SE, B240 Mayo, Box 292, Minneapolis, MN 55455. E-mail: kanal001@tc.umn.edu.

0897-1889/00/1302-1055$10.00/0
doi:10.1053/jdim.2000.6894

and significant errors ($P = .009$ and $P = .008$, respectively). The error rates for the male and female speaker groups were found to exhibit no statistically significant differences.

DISCUSSION AND CONCLUSIONS

The overall error rate in the current study was found to be 10.3% ± 3.3%. This compares with the 2.4% error rate reported by Herman for a discrete speech recognition system.[7] Differences in error definition and accounting between the two studies may contribute to the difference in reported accuracy. Taking the reported discrete speech results at face value, however, it appears that the convenience and efficiency of continuous speech has been included at the expense of an approximately fourfold increase in overall error rate.

The rate of significant errors was found to be 7.8% ± 3.4% (still approximately three times greater than the overall error rate previously reported for the discrete speech system). On average, the number of significant errors in an 87-word report (the observed mean word length of our test set of reports) would be about 7. The rate of subtle significant errors was computed to be 1.2% ± 1.6%. On average, the number of subtle significant errors in an 87-word report would be about 1.

Statistically significant differences in the accuracy were observed for the overall and significant errors as a function of the native English speaker status, although the error rates were fairly similar to one another. No statistically significant differences were seen between the male and female speaker groups.

Our evaluation methodology was useful for evaluating speech recognition accuracy, and was sensitive to subtle accuracy differences between groups of speakers. The speech recognition software is approximately 90% accurate, overall. Routine use of this system throughout a radiology practice is currently limited more by practical implementation issues than speech recognition accuracy. These practical implementation issues include the convenience of the existing transcription operation to the radiologists, the convenience of the computer-based transcription system, the current examination turnaround time to the clinical physicians, and the presence of other electronic systems such as RIS and PACS. However, applications in niche areas such as the emergency room may benefit from use of the system. It is expected that the use these systems interfaced with RIS and PACS will remove the major practical impediments to routine applications.

REFERENCES

1. Leeming BW, Porter D, Jackson JD, et al: Computerized radiologic reporting with voice data-entry. Radiology 138:585-588, 1981

2. Robbins AH, Horowitz DM, Srinivasan MK, et al: Speech-controlled generation of radiology reports. Radiology 164:569-573, 1987

3. Smith NT, Brien RA, Pettus DC, et al: Recognition accuracy with a voice-recognition system designed for anesthesia record keeping. J Clin Monit 6:299-306, 1990

4. Holbrook JA: Generating medical documentation through voice input: The emergency room. Top Health Rec Manage 12:49-57, 1992

5. Reed RA: Voice recognition for the radiology market. Top Health Rec Manage 12:58-63, 1992

6. Clark S: Implementation of voice recognition technology at Provenant Health Partners. J Am Health Inform Manage Assoc 65:34-38, 1994

7. Herman SJ: Accuracy of a voice-to-text personal dictation system in the generation of radiology reports. AJR 165:177-180, 1995

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