Voice Recognition Dictation: Radiologist as Transcriptionist

John A. Pezzullo,¹ Glenn A. Tung,¹ Jeffrey M. Rogg,¹ Lawrence M. Davis,¹ Jeffrey M. Brody,¹ and William W. Mayo-Smith¹

Continuous voice recognition dictation systems for radiology reporting provide a viable alternative to conventional transcription services with the promise of shorter report turnaround times and increased cost savings. While these benefits may be realized in academic institutions, it is unclear how voice recognition dictation impacts the private practice radiologist who is now faced with the additional task of transcription. In this article, we compare conventional transcription services with a commercially available voice recognition system with the following results: 1) Reports dictated with voice recognition took 50% longer to dictate despite being 24% shorter than those conventionally transcribed, 2) There were 5.1 errors per case, and 90% of all voice recognition dictations contained errors prior to report signoff while 10% of transcribed reports contained errors. 3). After signoff, 35% of VR reports still had errors. Additionally, cost savings using voice recognition systems in non-academic settings may not be realized. Based on average radiologist and transcription salaries, the additional time spent dictating with voice recognition costs an additional \$6.10 per case or \$76,250.00 yearly. The opportunity costs may be higher. Informally surveyed, all radiologists expressed dissatisfaction with voice recognition with feelings of frustration, and increased fatigue. In summary, in non-academic settings, utilizing radiologists as transcriptionists results in more error ridden radiology reports and increased costs compared with conventional transcription services.

KEY WORDS: Voice recognition dictation, radiologist, transcriptionist

INTRODUCTION

Computerized voice recognition (VR) for radiology reporting was first described in 1981,¹ and had remained more or less a novelty in the following two decades. Early systems were hampered by limited vocabularies and required 3 to 6 h training sessions before the system recognized the user's voice. Such systems only allowed discrete speech recognition, which proved to be the most serious disadvantage of the new technology. Discrete speech algorithms required a pause of 100 to 200 ms between each word substantially altering natural speech to a....series....of....starts....and....stops. Recent advances in computer processing speed and software have allowed continuous voice recognition to become a reality by more closely mimicking natural speech patterns. Specifically, voice recognition speech engines no longer recognize words as individual components of a sentence. Instead, language is packaged into a series of sounds called phonemes. The speech engine software

From the Department of Diagnostic Imaging, The Warren Alpert School of Medicine, Brown University, Providence, RI, USA.

Correspondence to: John A. Pezzullo, Department of Diagnostic Imaging, Rhode Island Hospital, 593 Eddy Street, Providence, RI, 02903, USA, Tel: +1-401-4445184; Fax: +1-401-4445017; e-mail: jpezzullo@lifespan.org

Copyright © 2007 by Society for Imaging Informatics in Medicine

Online publication 7 June 2007

doi: 10.1007/s10278-007-9039-2

then attempts to interpret phoneme segments using a statistical model (hidden Markov models) that encodes the probability that a given phoneme segment is correct based on the adjacent segments.² In practical terms, this means that the speech engine is more likely to recognize strings of words rather than individual words.

As such, continuous voice recognition dictation systems are becoming more popular by providing a viable alternative to conventional transcription services. An estimated 30% of radiology departments in the United States either have or plan to install voice recognition systems.³ Proponents of voice recognition point to marked reduction in report turnaround times, increased transcription savings, and relative ease of integration with existing radiology information (RIS) and picture archival and communication (PAC) systems. Detractors of such systems point to the fact that the task of transcription is being shifted to the radiologist who is not necessarily receiving any benefit of the cost savings, which are realized by the hospitals. In addition, more time is spent dictating and editing each report with diminished speech recognition accuracy compared to conventional transcription. As a result, radiologists may be less satisfied with VR systems compared to conventional transcription services.⁴

In this article, we describe our experience with a commercially available voice recognition dictation system compared to a conventional transcription service, both of which are employed in an outpatient practice setting. We specifically compared reports dictated with each system to determine (1) transcription times, accuracy, and report finalization times; (2) the costs associated with the use of each system. We also discuss the impact voice recognition transcription has had on the appearance and length of radiology reports and briefly discuss the impact of VR on user's satisfaction.

MATERIALS AND METHODS

Institutional review board approval was not required for this study. Our outpatient imaging center, opened in 2001, provides both computed tomography (CT) and magnetic resonance (MR) imaging services, and is a collaboration between a for-profit imaging company, the academic radiol-

ogy department, and the adjacent 715 bed academic institution. In addition, the center serves as the central MR interpretation site for four additional out-patient imaging offices in the outlying community and employs two full time radiologists who are responsible for the interpretation of approximately 75 MR and 25 CT studies daily. All imaging studies are sent and reviewed on a PACs workstation (Fuji Synapse, Tokyo Japan). All onsite MR and CT studies (~50 exams/day) are dictated using a commercially available continuous voice recognition software package (TalkTechnology version 2.1.28, AGFA Medical Systems, Germany) and a noise canceling microphone (Phillips, USA). The computer system employs a 2-GHz Pentium IV (Intel, Santa Clara, CA) PC with 640 MB RAM, 8 GB hard drive, and the Microsoft Windows XP Professional 2000 (Microsoft, Redmond, WA) operating system, and the appropriate network and sound cards.

All offsite MR imaging examinations are dictated using a conventional dictation station (Dictaphone, Lanier, Atlanta GA). MR imaging studies performed offsite are given a priority transcription code and are typically transcribed immediately once dictated. The transcribed reports are then faxed back to the imaging center for review and sign-off. Once signed, they are faxed back to the transcriptionists. All six full time transcriptionists are employed by a private corporation. Both the transcription pool and the continuous voice recognition system are connected via a local area network to radiology information systems (IDXrad version 9.0; IDXrad, Burlington VT for offsite studies, and Costar version 2.1; Clearview software, Amherst NH for onsite studies). Once transcribed by either method, all radiology reports are faxed, unless specifically requested by the referring physicians to be sent by conventional mail. Emergent findings are immediately called. Physicians who are given access to the radiology information systems may also view written reports immediately once transcribed via the worldwide web.

STUDY DESIGN

Between June and August of 2004, we prospectively compared continuous speech recognition and traditional transcription dictation. One hun-

dred consecutive MR imaging examinations of the cervical or lumbar spine were dictated using continuous voice recognition, while 100 different consecutive MR imaging examinations, similar to the cervical and lumbar spine performed at the off-site centers, were dictated using conventional transcription services. All examinations were performed for routine indications of low back pain or radicular symptoms. No postoperative studies or exams requiring comparison with another imaging test were included in the analysis. All examinations were interpreted at the same workstation by seven attending radiologists with an average of 3 years of experience using the voice recognition system. All seven radiologists speak English as the native language without speech impediments, and all underwent the same initial voice recognition training and enrollment process. The interpreting radiologists was asked to first review the study, then to dictate the examination on either system, noting the start and stop times. The voice recognition dictation time included editing time and the number of errors noted. Types of errors included words not recognized or misrecognized, punctuation and spacing errors, and formatting errors. A word or formatting command was considered erroneous if it was not transcribed as intended. The conventional transcription dictation time was determined by the dictation start and stop time and the correction time once transcribed. Voice recognition macros or templates were not allowed for purposes of the study. Additional data recorded for each examination included number of words per report including formatting commands and punctuation, number of errors before and after each report was signed, report finalization time [time report is available to the referring clinician determined on the radiology information system (RIS) system]. Report finalization times for those reports dictated with voice recognition include the dictation time plus the correction and report review/sign-off time. For those reports dictated conventionally, report finalization times include dictation time plus transcription time plus review and fax times. In addition, each report was reviewed to determine the percentage of finalized reports containing errors between each dictation system. Subjective information from each of the seven radiologist participants were also collected and are presented in the analysis.

Cost Analysis

In fiscal year 2003, the six full time employees in the transcription department transcribed 122,590 reports. Total costs in the department including salary, benefits, and service contracts were \$344,571. The average total costs-per-dictation was \$2.97. Based on a 40-h workweek and 50 workweeks/year, the average hourly wage is calculated as \$175/h (\$2.92/min) for a radiologist and \$16.00/h (\$0.27 /min) for the transcriptionists. At the outpatient imaging center, an average of 100 exams are dictated per day and are approximately evenly split between those conventionally transcribed and those dictated using the VR system.

RESULTS

Patient Demographics

Sixty-four MR examinations of the lumbar spine and 36 MR examinations of the cervical spine were conventionally transcribed, while 59 MR examinations of the lumbar spine and 41 MR examinations of the cervical spine were transcribed using VR. The average patient age in the conventional transcription group was 57 years (range: 19–87 years) and 55 years (range: 21–84 years) in the VR group.

Report Differences

MR reports of the cervical and lumbar spines using the voice recognition dictation system took approximately 50% longer to dictate and tended to be 24% shorter in length than those transcribed conventionally. The average dictation time for VR reports was 4.11 min (SD 2.11 min) versus 2.02 min (SD 1.23 min) ($p < 0.0001$) resulting in a time difference of 2.09 min (correction time). The correction time for transcribed reports was 10 s and is included in the average dictation time. Reports transcribed conventionally were essentially error free (0.12 error/report), while those transcribed using the voice recognition system had 5.1 errors/report ($p < 0.0001$). Overall, recognition accuracy was close to 100% for transcribed reports, while voice recognition accuracy was 96%. In addition, 89% (89/100) of VR reports contained an error before report sign-off, while only 10% (10/100) of conventionally transcribed

reports contained errors. Interestingly, 35 (35%) of VR reports re-reviewed after sign-off contained errors. Only three transcribed reports contained errors (3%). Types of errors included spacing and punctuation errors or misrecognized words not initially identified before sign-off. None of the errors were considered significant to alter patient management. There were no spelling errors. The average report finalization time was 33.5 min for VR reports and 72.3 min ($p<0.001$) for transcribed reports. Results are summarized in Table 1.

Costs

Based on the average hourly radiologists wage and the mean correction time per case, dictation with voice recognition results in added costs of \$6.10/case. This amounts to a daily cost of \$305.00, a weekly cost of \$1,525.00 (assuming five working days per week), and a yearly cost of \$76,250.00. As stated, the average total cost per dictation conventionally transcribed in our system is \$2.97. Based on the average transcriptionists salary, the additional correction time would cost an additional \$0.56/case. If we assume that all VR cases dictated per day are routine cervical and lumbar spines, then, an additional 104.5 min is added in correction time per day. If this additional time were spent dictating ten routine examinations of the spine, with conservative global reimbursement of \$400.00 per case, an additional 2,500 MR examinations could be read per year, equating to \$1,000,000 in additional revenues annually.

DISCUSSION

Proponents of continuous voice recognition dictation point to marked transcription cost sav-

ings and decreased report turnaround times as the main benefits of such systems with the potential to save hundreds of thousands of dollars per year in transcription costs and provide near instantaneous report availability.^{3,5-9} What is unclear is the impact voice recognition has had on radiologists who are now faced with the additional responsibility (both physical and financial) of providing transcription services.

Our results suggest that radiologists are not good transcriptionists. Routine cervical and lumbar MR reports took an average of 50% longer to dictate despite being 24% shorter in length. A full 90% of reports contained errors before sign-off, and remarkably, 35% still had errors *after* sign-off. Only three of the conventionally transcribed reports had errors after sign-off. While none of the errors would be considered significant to alter patient management, some would argue that a less than perfect radiology report is unacceptable given that the report is the primary means of communication between the radiologist and the referring community. Malpractice attorneys will often note typos and grammatical errors in reports as examples of a radiologists carelessness or disinterest.¹⁰ The lack of errors in conventionally transcribed reports supports the contention that transcriptionists do more than transcribe recorded words; they may, in fact, be over-reading each case based on the dictating physician's habits and style.¹¹

Voice recognition reports tend to be substantially shorter in length than those transcribed conventionally. Ramaswamy et al.⁹ reported a 37% reduction in MR report length utilizing voice recognition presumably as a compensatory measure for added correction time. This fact supports the contention that radiologists will alter their dictation style when using VR dictation.^{4,12} It is

Table 1. Summary of Results Comparing Voice Recognition Dictation and Conventional Transcription Services

	Transcription	Voice Recognition
Reports (N)	100	100
Total Dictation Time	2.02 min (SD 1.23)	4.11 min (SD 2.11 min)
Words per report	225.6 (SD 86.2)	181.7 (SD 62.2)
Errors per report	0.12 (SD 0.36)	5.1 (SD 4.4)
Reports with errors	10 (10%)	89 (89%)
Finalized reports with errors	3 (3%)	35 (35%)

SD Standard deviation

unclear what this means in terms of quality of the radiology report, and the amount of information contained in the report. There was a noticeable difference in dictation style among several of our radiologists between the dictation systems. While the VR reports tended to be shorter, they contained more incomplete or fragmented sentences. A typical example of this dictation anomaly would be: *at C4–5, no discs or canal stenosis*. This same sentence conventionally transcribed read: *at the C4–5 level, there was no evidence of focal disc herniation or central canal stenosis*.

Voice recognition reports also contain more errors than those conventionally transcribed with reported recognition rates between 90 and 95%.^{5,8,9,13} We report an overall recognition rate of 96%, slightly greater than that reported. The likely reason for the difference is that only routine studies were dictated and tend to be shorter than the more complex studies normally performed at our site. The recognition rate is likely to be even higher if macros are used. A macro is a standardized report typically describing normal or frequent findings. For the purposes of this study, macros were not employed. While we recognize that the uniform use of macros will substantially decrease dictation times and errors, most examinations performed at this outpatient center, due to complexity, do not routinely fall into a macro-report format. It is also likely that error rates and dictation times will diminish with newer versions of speech recognition software.

The costs associated with using radiologists as transcriptionists may be prohibitive. Issenman and Jaffer¹³ report a 100% increase in dictation costs using a voice recognition system compared with conventional transcription. Our results are similar. The additional time requirement to correct each VR case in our analysis costs an additional \$6.10 per case based on the radiologists salary and adds 105 min to each day. The *total* transcription costs (excluding start-up costs) per case in our office practice is \$2.97. The additional time added per day is likely an underestimation, as the majority of cases dictated using VR are complex follow-up MR or CT cases as the center serves as the outpatient imaging center for the adjacent tertiary care hospital. The opportunity costs may be even higher, as the “lost” revenue from dictating with VR may be substantial. The use of VR has raised workflow and manpower issues within our prac-

tice. It is likely that the outpatient center could be staffed with 1.5 full time employees (FTE) if VR was not in use, or that additional MR imaging studies could instead be interpreted onsite. What about the cost savings as reported in the literature and promised by the vendors? While these cost savings may be beneficial to the institutions that have embraced voice recognition technology, it is unclear if any of the saved revenue has been transferred back to the radiologists either as salary support or additional departmental capital.

The promise of decreased report turnaround time, while realized in the hospital setting, may not be as significant in the outpatient environment. As reported in our study, finalization times for VR reports were approximately half those of conventionally transcribed reports, despite the priority status given to MR reports conventionally transcribed. While this number is statistically significant, in practical terms, it likely means little as nearly all of the reports are received within the referring clinicians office within 2 h after dictation and sign-off, sooner if the physician has the ability to view the reports and images on line. Positive results are always called to the referring MD’s office. We do recognize that report turn around time is an issue in a hospital setting as studies are typically dictated by fellows or residents before being signed by an attending radiologist.

Finally, radiologists may not be happy transcriptionists. In a departmental survey performed at Duke University, 77% of respondents felt that voice recognition increased the daily workload, and 42% of respondents felt the system was poor or a waste of time.⁴ All seven radiologists in our study have expressed some degree of dissatisfaction with voice recognition. Common sentiments expressed included increased user fatigue and increased frustration in addition to alteration of dictation styles and substance.

SUMMARY

Voice recognition dictation systems result in greater physician time, physician dissatisfaction, and have more errors than conventional transcription. In addition, voice recognition has greater aggregate costs when radiologists time for correction is incorporated.

ACKNOWLEDGEMENTS

Many thanks to Jennifer Loring, PACs coordinator, for help with data collection.

REFERENCES

1. Leeming BW, Porter D, Jackson JD, Bleich HL, Simon M: Computerized radiologic reporting with voice data-entry. *Radiology* 138:585-588, 1981
2. Zafar A, Overhage JM, McDonald CJ: Continuous speech recognition for clinicians. *J Am Med Inform Assoc* 6:195-204, 1999
3. Mehta A, McCloud TC: Voice recognition. *J Thor Imag* 18:178-182
4. Pappas JN, Bissett GS, Ravin CE: show Satisfaction survey from 77 radiologists in a single large academic radiology department using a voice recognition dictation system—a one-year experience. *Radiology* 218:611, 2001
5. Robbins AH, Horowitz Dm, Srinivasan MK, et al: Speech-controlled generation of radiology reports. *Radiology* 164:569-573, 1987
6. Langer S: Radiology show speech recognition: work-flow, integration, and productivity issues. *Curr Probl Diagn Radiol*. 31:95-104, 2002
7. Mehta A, Dreyer K, Boland G, Frank M. Do picture archiving and communication systems improve report turnaround times. *J Digit Imaging* 13(2 suppl 1):105-107, 2000
8. Rosenthal DI, Chew SS, Dupuy DE, et al: Computer based speech recognition as a replacement for medical transcription. *AJR* 170:23-25, 1998
9. Ramaswamy MR, show Chaljub G, Esch O, fanning DD, vanSonnenberg E: Continuous speech recognition in MR imaging reporting: advantages, disadvantages, and impact. *AJR* 174:617-622, 2000
10. Smith JS, Berlin L: Signing a colleague's radiology report. *AJR*:27-30, 2001
11. Schwartz LH, Kijewski P, Hertogen H, Roosin PS, Castellino RA: Voice recognition in radiology reporting. *AJR* 169:27-29, 1997
12. Hansen GC, Falkenbach KH, Yaghamai I: Voice recognition system (letter). *Radiology* 169:580, 1988
13. Issneman RM, Jaffer IH: Use of voice recognition software in an outpatient pediatric specialty practice. *Pediatrics* 114:290-293, 2004