

Learning Curve of Speech Recognition

Tomi A. Kauppinen · Johanna Kaipio · Mika P. Koivikko

Published online: 19 June 2013
© Society for Imaging Informatics in Medicine 2013

Abstract Speech recognition (SR) speeds patient care processes by reducing report turnaround times. However, concerns have emerged about prolonged training and an added secretarial burden for radiologists. We assessed how much proofing radiologists who have years of experience with SR and radiologists new to SR must perform, and estimated how quickly the new users become as skilled as the experienced users. We studied SR log entries for 0.25 million reports from 154 radiologists and after careful exclusions, defined a group of 11 experienced radiologists and 71 radiologists new to SR (24,833 and 122,093 reports, respectively). Data were analyzed for sound file and report lengths, character-based error rates, and words unknown to the SR's dictionary. Experienced radiologists corrected 6 characters for each report and for new users, 11. Some users presented a very unfavorable learning curve, with error rates not declining as expected. New users' reports were longer, and data for the experienced users indicates that their reports, initially equally lengthy, shortened over a period of several years. For most radiologists, only minor corrections of dictated reports were necessary. While new users adopted SR quickly, with a subset outperforming experienced users from the start, identification of users struggling with SR will help facilitate troubleshooting and support.

Keywords Speech recognition · Radiology reporting · Workflow · Statistic analysis

Introduction

Integrated picture archiving and communication systems (PACS) and radiology information systems (RIS) can potentially provide both fast and location-independent image distribution and reporting to hospitals and outpatient clinics alike. The reporting process can, however, spoil the equation. Transportation of digital tapes, transcription, and the radiologists' correction of and signoff on transcripts are all stages of the process during which queues or staff absences can result in delays measured in days rather than hours. Delays in the radiologists' finalization of reports may be particularly detrimental for report turnaround times (RTT) [1, 2].

From the current literature, the benefits from computerized speech recognition (SR), its immediate finalization, and the resultant rapid reporting, are indisputable [1, 2]. Accumulating evidence shows that SR significantly speeds RTT [3, 4]. Thus, SR accelerates the complete patient care process, with emergency rooms typically benefiting most from the streamlined process.

Any required proofing by radiologists raises, however, distrust of SR [5] and is, in addition to inevitable resistance to changing work habits [6], probably the main reason why SR has not been adopted more quickly; however, if the importance of SR is evaluated process-wise [1] from a complete hospital's or hospital district's point of view, it makes sense. Incentives to radiologists for using SR may even be appropriate. The radiologist's burden from SR is, however, difficult to measure objectively. Scientific data on how fast SR can be adopted by radiologists and how skilled the users will become has been, to date, unexplored.

The purpose of this study was to analyze the SR reporting of experienced and new users, focusing on number of corrections and report lengths, and, by using a longitudinal measurement, estimate the learning curve of a new user who takes up SR.

T. A. Kauppinen (✉) · J. Kaipio · M. P. Koivikko
HUS Medical Imaging Center, Helsinki University Central
Hospital, PO Box 750, 00029 Helsinki, Finland
e-mail: tomi.kauppinen@hus.fi

Materials and Methods

HUS Medical Imaging Center comprises 31 radiology departments and produces imaging services throughout the hospital district of the Helsinki and Uusimaa region (Finland), serving 1.5 million inhabitants. Studies include regular x-ray, angiography, ultrasound and CT, as well as MRI, totaling to 1 million imaging studies annually.

SR has been used in the HUS Medical Imaging Center since 2005 [3]. From the start, 13 radiologists used SR since 2005. The current SR (SpeechMagic 6.1 SP2, Nuance Communications Inc., Burlington, MA, USA) was expanded to all our departments in 2010, adding 104 totally new users. Currently, we have 154 radiologists in 31 radiology departments using 146 PACS workstations with SR capability and online editing available during or at the end of dictation.

The study sample analyzed consisted of SR log files from March 11, 2010 to August 6, 2011, a total of 0.25 million SR reports. First, we identified a reference group of experienced users, including 11 radiologists (mean age \pm SD, 44.4 ± 9.7 , eight males and three females) who had adopted SR since 2005 (mean 4.5 years of experience). The same radiologists had participated in our 2008 study of report turnaround times [3]. Secondly, after exclusion of all residents (who regularly switch departments and are less experienced in radiology), all radiologists who switched departments during the sample or whose work history was unreliably documented, and all users with any previous SR experience, 71 radiologists (mean age \pm SD, 48.5 ± 7.1 , 26 males and 45 females) were approved as the new-users group. The radiologists in the group of experienced radiologists were not more senior radiologists than the new users of SR. The study profile of both groups ranged from simple radiographs (with short reports) to MRIs and CTs, without any significant shift of amount of studies during the samples.

No reporting templates were used. At the start, each new user received training from our PACS help-desk staff, where new users dictated standard text (approximately 10 min) to the system. Then, users adjusted the audio settings for recording volume and silence detection level. The new users created 156,046 SR reports during the study period, and the 11-member reference group 45,001. Because of different SR roll-out times for each department, the actual starting dates and corresponding follow-up times for the new users differed. Therefore, for each new user, we identified from the data his/her first day of usage and recalculated the data in order to obtain time series statistics for the first day, second day (and thereafter) of SR usage. To ensure reliable statistics, a cutoff at 41 weeks of SR use was determined, at which point, data were still available for 90 % of users. During their 41 tracked weeks, the new users dictated a total of 122,093 reports. Because the reference group of experienced users had completed their learning curve years before, not during

the study period, their statistics were recorded and averaged from the 41 weeks starting from March 2010—24,833 reports in total.

We analyzed radiological reports done by SR and then compared the results between new and experienced users to find differences in error rates (correction rate, number of corrections in keystrokes/report) and unknown words. Because their report lengths appeared unexpectedly different, an analysis of report lengths was also conducted.

To guarantee data consistency and validity, the SR's logging functionality was exhaustively tested by an experienced radiologist who dictated test reports using variable sound file lengths, pauses, and word- and character-counts; different patterns of correction; and varying numbers of non-existent words intentionally absent from the system's vocabulary. To correlate the error rate and the number of reports, a nonparametric correlation (Spearman test, 2-tailed p value) was calculated.

Results

New users' report length was relatively constant, with a mean of 433 ± 375 characters (mean \pm SD), throughout the 41-week period (Fig. 1). Experienced users created reports that were 31 % shorter, with a mean 298 ± 263 characters (mean \pm SD). New and experienced users' sound file lengths similarly differed: 52.0 ± 51.8 versus 37.2 ± 40.3 s (mean \pm SD) (Fig. 2). To better understand these differences in report lengths and to illuminate possible evolution over a longer period, we investigated experienced users' earlier history of report lengths. Measurements from their department from 2007 and 2011, 2 and 6 years after SR adoption showed at 2 years a mean report length of 559 characters (sample of 34,684 reports from 12 radiologists) and at 6 years, 308 characters (45,001 reports from 11 radiologists).

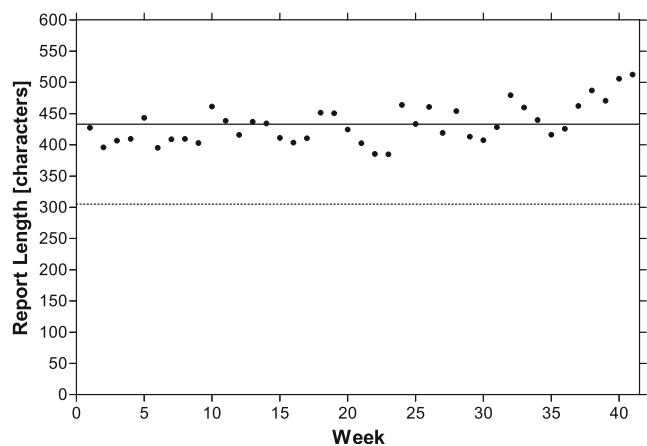


Fig. 1 Mean report length after SR adoption. New users' weekly averages (dots), average of the complete follow-up for new (solid line) and experienced (dotted line) users, respectively

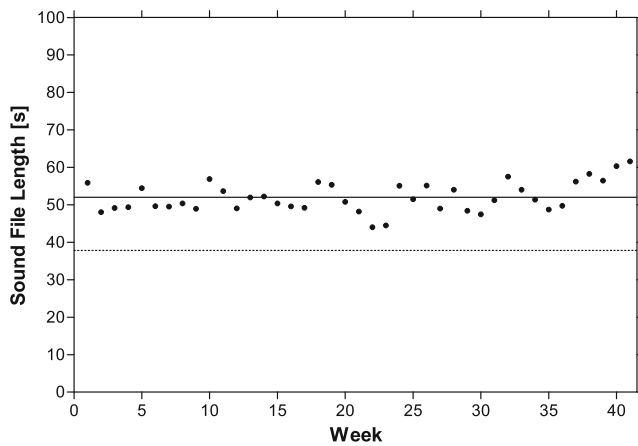


Fig. 2 Mean sound file length after SR adoption. New users' weekly averages (dots) and their 41-week average (solid line), and experienced users' average (dotted line)

New users at the beginning of SR had 77 % more unknown words (0.46 per report) than did the experienced users (0.26). However, this number declined rapidly, with a plateau of 0.36 (38 % more than the experienced users) achieved after approximately 10 weeks (Fig. 3). The time series of new users' character-based error rates, i.e., number of characters corrected prior to finalization, is presented in Fig. 4. At the start, the new users corrected 15 characters (of 433); experienced users 6 (of 298) per report. The new users' error rate was, allowing for differences in report length, 60 % higher at the start (3.4 versus 2.1 %), decreased more rapidly in the beginning (30 % higher at 14 weeks), and improved throughout the following weeks (14 % higher at 41 weeks). A subset of users succeeded well from the beginning, whereas some did worse (Fig. 5), without differences in demographics or radiology experience. A significant correlation ($p < 0.0001$) exists between users' error rates and their cumulative number of reports (Fig. 6).

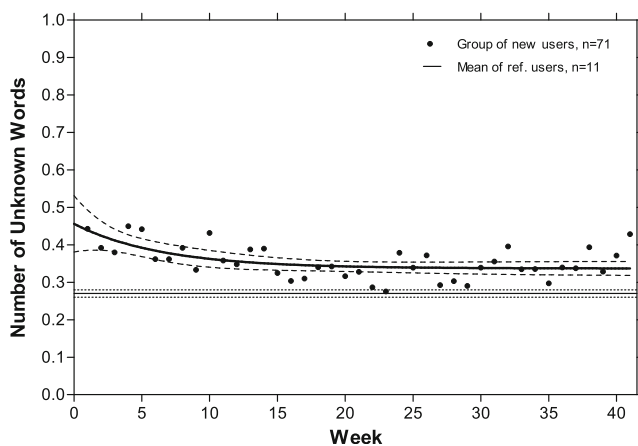


Fig. 3 Occurrence of unknown words per report during the follow-up. Weekly average, mean curve, and 95 % CI for new and experienced users

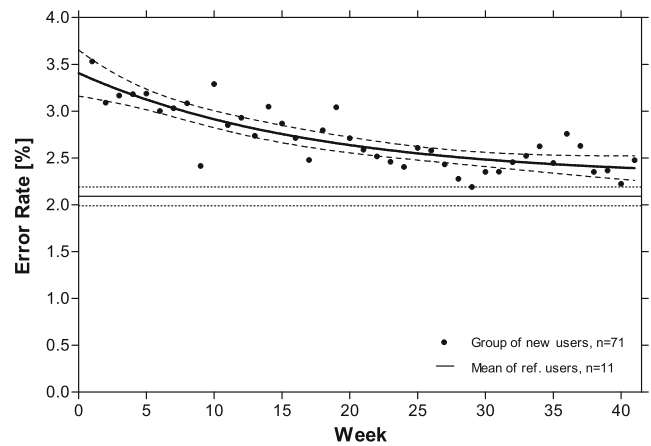


Fig. 4 Time series of character-based error rates (percent of characters). New users' weekly average, mean curve, and 95 % CI for new and experienced users

Discussion

According to our results, radiologists experienced with SR reports need only minimal editing and new users, on average, quickly become nearly as skilled, with error rates differing by no more than 50 % after 4 weeks of SR experience. Some users presented, however, a very unfavorable learning curve, literally struggling with SR. Whether their trouble arose from inadequate training or malfunctioning hardware or software remains unknown, but our methodology certainly identified those needing help.

Our study was planned after the reports were made, eliminating any potential influence on radiologists' behavior. Data was validated consistent with actions performed on a workstation. However, applicability to the non-Finnish world deserves consideration. Technically, Finnish is hard for SR [3], because of its core vocabulary and agglutinated derivative suffixes that result in millions of words. Because agglutinated Finnish words tend to be long, number of

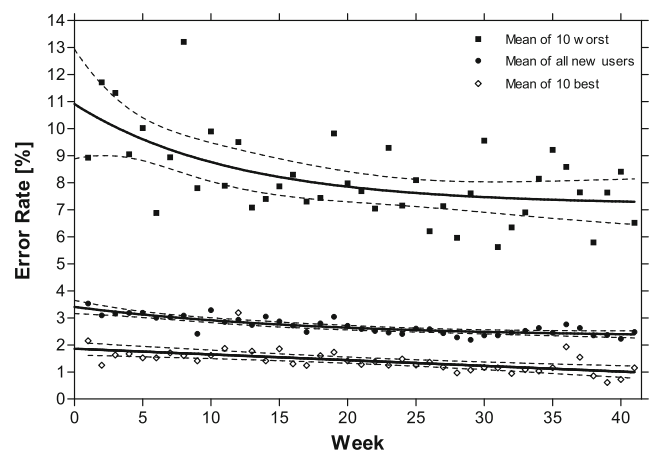


Fig. 5 Weekly error rates and mean curves with 95 % CI for all 71 new users, and of them, the 10 best and 10 worst succeeding

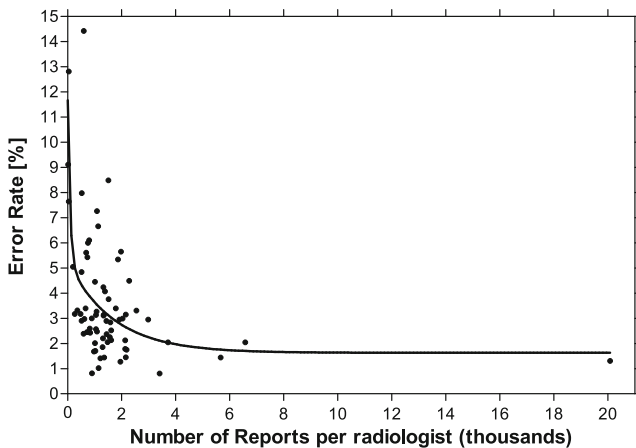


Fig. 6 Correlation of each user's error rate and cumulative number of reports after 41 weeks of follow-up, and two-phase exponential decay (*curve*) of the correlation

corrected characters was considered more accurate than number of words. Words unknown to SR occur frequently and likely raise our error rates. Most new words have only a suffix, with two to three characters to edit. The estimated effect of Finnish was therefore minor.

Low-correction rates are not necessarily synonymous to efficiency: an antithesis could be that efficient radiologists report quickly and exert many corrections. Our data, however, suggests a correlation of low-correction rates and productivity (Fig. 6). We also acknowledge that different generations of SR software probably possess different levels of efficiency in iterative learning. A controlled study comparing them would be burdensome and even become obsolete before publication. We think, however, that our results are valid for the near future, helping those deciding whether to omit transcription and implement SR.

SR's shorter RTT offers economic and medical advantages. In one study of 21,595 reports, RTT fell 81 %, from 24.8 to 4.4 h [3]. One study of 305,892 proof read and finalized reports [4] confirmed concurrent with a shift from digital dictation to SR, a reduction from 28 h to 12.7. Both studies reported concurrent productivity increases [3, 4].

With any method to produce text, errors occur. Chang et al. [7] studied SR errors that escaped the radiologists' proofreading. In 990 reports by 19 radiologists, each with several years of SR experience, 6 % of x-ray reports had errors, but other modalities had 3.5-fold. They also reported radiologists' error rates variable and postulated that "causes may range from pronunciation, clarity, and speed of the radiologist dictating the report, to failure to proofread the reports accurately. Carelessness of the reporter may also be a possibility." A study of 1,160 SR and 727 transcribed reports found errors in 4.8 and 2.1 % concluding that errors are related to a noisy environment, high workload, and radiologists speaking English as non-native speakers. Of their 71

erroneous reports, irrespective of the dictation method used, 52 % had errors that affected understanding of the report, but with none considered to adversely affect patient care [8]. One group [2] noted importantly, that typists also perform unequally. Another [9], with post-SR statistical error detection, found an error reduction of 96 %. These studies [7–9] certainly justify further research into radiology reporting—both of the occurrence and significance of errors resulting from different tools.

Our new users' reports were consistently longer than those by the experienced users. To understand this difference, we sought old data for the experienced users and found their reports 45 % longer in 2007, a phenomenon previously undocumented in the literature. Our perception is that SR facilitates more focused, structurally efficient reports, and a development occurring not over months but over years. Antithetically, SR causes radiologists to report more briefly, because editing could be burdensome. Further study is vital to assess SR's long-term impact on information content and readability; and how much SR adjusts to radiologists and how much radiologists' language adjusts to SR usage.

Successful adoption of SR requires participation. Krishnaraj et al. [4] associated users' short RTT with openness to learning SR's features, in addition to promptness in checking the report queues. Those with the greatest RTT reductions taught the SR vocabulary more often, contacted IT support more, and consistently used SR to improve their workflow. Radiologists revising trainee reports at the time of image review benefited most [4]. We consider SR report proofreading straightforward and more fail-safe when performed with the images still memorized and reviewable. For comparison, finalization of transcripts deserves some attention. One study found 14 specific reasons for signoff delays [10]. By carefully scrutinizing the processes, some of these obstacles can be overcome [11], and we believe that process-oriented RIS and PACS address some of them. However, some delays, especially those related to radiologists' availability, persist. SR's immediate finalization simplifies the complete process.

Undoubtedly, some radiologists still consider proofing required by SR to be a burden. In most Finnish hospitals, no signoff is required for radiology transcripts, which is certainly convenient but also hazardous. Appropriate incentives could be used to encourage SR usage when advantageous; for instance, additional earnings for emergency room examinations reported using SR. We acknowledge that such incentives are impossible in our public health care system with its monthly salaries and minimal liberty to offer bonuses, but it could be a viable solution for hospitals with examination-based reimbursements. SR macros and templates, and tightly integrated workflow—beyond the scope of this study—may enhance radiologists' productivity,

especially in normal findings, which could make a sufficient incentive alone.

In conclusion, our study of experienced and new SR users shows that, on average, radiologists exert only minimal proofreading thus relieving concerns as to radiologists' doing secretarial work. Experienced users altered only 6 and new users 11 characters to finalize a report. Radiologists pick up SR quickly, but also remarkably differently: Some outperform the experienced from the start, while most experience a learning curve of several months. The reports tend to shorten over a span of years. A subset of new users may require special attention to track and solve technical issues or user errors—a process that could be proactively triggered by monitoring error rates and should be carefully planned for new and existing SR installations alike.

Acknowledgments This work was partly supported by the special governmental subsidy for health sciences research.

References

1. Boland GW, Guimaraes AS, Mueller PR: Radiology report turnaround: expectations and solutions. *Eur Radiol* 18:1326–1328, 2008
2. Hart JL, McBride A, Blunt D, et al: Immediate and sustained benefits of a “total” implementation of speech recognition reporting. *Br J Radiol* 83:424–427, 2010
3. Kauppinen T, Koivikko MP, Ahovuo J: Improvement of report workflow and productivity using speech recognition—a follow-up study. *J Digit Imaging* 21:378–382, 2008
4. Krishnaraj A, Lee JKT, Laws SA, et al: Voice recognition software: effect on radiology report turnaround time at an academic medical center. *Am J Roentgenol* 195:94–197, 2010
5. Chen JY, Lakhani P, Safdar MN, et al: Letter to the editor: voice recognition dictation: radiologist as transcriptionist and improvement of report workflow and productivity using speech recognition—a follow-up study. *J Digit Imaging* 22:560–561, 2009
6. Bramson RT, Bramson RA: Overcoming obstacles to work-changing technology such as PACS and voice recognition. *Am J Roentgenol* 184:1727–1730, 2005
7. Chang CA, Strahan R, Jolley D: Non-clinical errors using voice recognition dictation software for radiology reports: a retrospective audit. *J Digit Imaging* 24:724–728, 2011
8. McGurk S, Brauer K, Macfarlane TV, et al: The effect of voice recognition software on comparative error rates in radiology reports. *Br J Radiol* 81:767–770, 2008
9. Voll K, Atkins S, Forster B: Improving the utility of speech recognition through error detection. *J Digit Imaging* 21:371–377, 2008
10. Seltzer SE, Kelly P, Adams DF, et al: Expediting the turnaround of radiology reports: use of total quality management to facilitate radiologists' report signing. *Am J Roentgenol* 162:775–781, 1994
11. Seltzer SE, Kelly P, Adams DF, et al: Expediting the turnaround of radiology reports in a teaching hospital setting. *Am J Roentgenol* 168:889–893, 1997