How to Cite:
Alonayzan, H. H. M., Alenezi, Talal S. S., Alshammari, Khalaf S. F., Albakr, Mohammed S. B., Alshammari, Sanad H. S., Alghadeer, Saleh O. A., … Alanzi, A. S. B. (2019). The potential of voice recognition technology in medical record documentation: Review. *Tennessee Research International of Social Sciences*, *1*(2), 7–16. Retrieved from https://triss.org/index.php/journal/article/view/42

The potential of voice recognition technology in medical record documentation: Review

Hamad Hassan Mohammed Alonayzan
KSA, National Guard Health Affairs
Email: Alonizanha@ngha.med.sa

Talal Sanian Salem Alenezi
KSA, National Guard Health Affairs
Email: Alenezyta@ngha.med.sa

Khalaf Saud Faryhan Alshammari
KSA, National Guard Health Affairs
Email: alshammarykh@ngha.med.sa

Mohammed Saad Bakr Albakr
KSA, National Guard Health Affairs
Email: al-bakrmo@ngha.med.sa

Sanad Hamdan Sanad Alshammari
KSA, National Guard Health Affairs
Email: aIshammarisa5@ngha.med.sa

Saleh Obaid Abdullah Alghadeer
KSA, National Guard Health Affairs
Email: Alghadersa@ngha.med.sa

Nezar Mohammad Mutlaq Alshammari
KSA, National Guard Health Affairs
Email: Alshammaryne@ngha.med.sa

Fahad Khalifah Salem Almughamis
KSA, National Guard Health Affairs
Email: almgamasfa@ngha.med.sa

Nuri Rawafa Alanzi
KSA, National Guard Health Affairs
Email: alenazinu@ngha.med.sa

ISSN: 2766-7464 (Online)
Publisher: Smoky Mountain Publishing
Manuscript submitted: 09 Sept 2019, Manuscript revised: 18 Oct 2019, Accepted for publication: 27 Nov 2019

Abdullah Ibrahim Hamran
KSA, National Guard Health Affairs
Email: Hamranab@ngha.med.sa

Fawaz Ayed Al-Sharari
KSA, National Guard Health Affairs

Ahmed Turki Alotaibi
KSA, National Guard Health Affairs
Email: A7mdoh85@gmail.com

Awad Shehab B Alanzi
KSA, National Guard Health Affairs

***Abstract*---**Background: Speech recognition (SR) systems have been used in medical reporting for over 20 years, converting spoken words into written text and allowing voice commands. Initially hampered by underdeveloped technology and unsatisfactory recognition error rates, progress has been made in algorithm design, system performance, and technology, with newer approaches emerging. Aim of Work: The objective of this study is to examine existing literature that evaluates the effects of speech recognition (SR) technology on clinical documentation. Methods: Studies conducted before December 2014 that reported clinical documentation employing SR were located by searching Scopus, Compendex and Inspect, PubMed, and Google Scholar. The examined outcome variables consisted of dictation and editing time, document turnaround time (TAT), speech recognition accuracy, mistake rates per document, and economic gain. Results: The majority of research focused on comparing speech recognition (SR) to dictation and transcription (DT) in the field of radiology. There was a significant level of variation across the studies. Document editing time was shown to be significantly longer when utilizing speech recognition (SR) compared to traditional typing (DT) in four out of six tests, with an increase ranging from 1876.47% to -16.50%. The duration of dictation rose in three out of five investigations, ranging from a 91.60% rise to a 25.00% decrease. The TAT showed a continuous increase while using SR instead of DT, increasing from 16.41% to 82.34%. Across all investigations, the improvement rate was 0.90% each year. The accuracy of speech recognition (SR) was documented in 10 investigations, ranging from 88.90% to 96.00%. It was observed that the accuracy of SR increased by 0.03% year as the technology advanced. The average number of mistakes per report significantly increased while employing SR, ranging from 0.05 to 6.66, in comparison to DT, which ranged from 0.02 to 0.40. The reporting of economic gains was inadequate. Conclusion: The findings indicate that SR is consistently developing and has some benefits for clinical documentation. Nevertheless, the data endorsing the use of SR is lacking in strength, and more inquiry is necessary to evaluate the influence of SR on the kinds, frequencies, and clinical results of documentation errors.

***Keywords*---**Health Record Systems, Electronic Health Records, Patient Care, Review, Ethical Matters.

Introduction

Commercially accessible speech recognition (SR) systems for medical reporting have been in existence for more than twenty years**.** [1] Speech recognition (SR) is a tool that helps with clinical documentation by converting spoken words into written text or by allowing users to manage interface functionalities via voice commands. Speech recognition (SR) has been effectively used in clinical environments, particularly for the transcription of radiology reports.[2,3] It is often employed with the radiological information system or image archiving and communication systems. Nevertheless, the use of SR has not been consistent across all therapeutic fields. On the other hand, SR is now extensively used in many consumer applications, such as interface control and question answering apps in smart phones. [4]

The first use of SR-based documentation was impeded by underdeveloped technology and recognition error rates that were deemed unsatisfactory in a clinical setting. However, significant progress has been achieved in the last two decades via continuous improvements in algorithm design and system performance.[5,6] The technology behind SR systems has seen significant advancements, particularly in the SR engines responsible for speech recognition, as well as the processing speed and memory capacity of the hardware used to handle speech input. In the past, early speech recognition systems used formal language grammars. However, they were replaced by probabilistic techniques like Hidden Markov models, which are being used today. Language models of this kind need the use of signal processing techniques to extract fundamental characteristics from speech data.[7-10] Additionally, statistical acoustic models are employed to describe the various sounds or phonemes found in speech. The development of SR techniques is ongoing, with newer approaches including structured speech and language models, conditional random fields, and maximum entropy Markov models. [11]

Aim of Work

The aim of this study is to provide a concise overview of the scientific literature that discusses the advantages and potential drawbacks of using speech recognition (SR) technologies for clinical recording purposes. The secondary objective is to investigate if the performance of speech recognition (SR) in clinical documentation tasks has shown improvement over time as this technology has become more advanced.

Methods

Article searches were conducted using Scopus, Compendex, Inspect, PubMed, and Google Scholar without any limitations on the publication date. The search query employed was: "speech recognition" or "voice recognition" or "Dragon Naturally Speaking" (the inclusion of other brand names did not yield any additional outcomes) and "medical record*" or "health record*" or "patient record*" or "nursing record*" or "clinical record*" or "radiolog*".

Speech Recognition Efficiency

Several studies examined the effects of speech recognition (SR) on the speed at which clinicians write or alter clinical documents, as well as its influence on the time it takes for organizations to process these documents.[12,14] Five researches documented the duration of dictation, which refers to the time required to generate a new clinical document using either speech recognition (SR) or direct transcription (DT) methods, as shown by studies [14–18]. Among these studies, four focused on analyzing radiology reports, while one examined general clinical notes. Out of the 18 participants, three had a rise in dictation time while utilizing speech recognition technology, increasing from 35.64% to 91.60%. Two studies shown that adopting speech recognition (SR) technology resulted in a decrease in dictation time compared to traditional typing (DT) by 10.87% and 25.00% respectively. Each research used a distinct speech recognition (SR) system, namely Philips Nuance, LTI, AGFA, and ASR Medispeak [14,16,17,18]. The sample sizes varied from a single subject, with a minimum of 18, to a whole department, with a maximum of 17. The length of the studies ranged from 4 weeks, with a minimum of 17, to 3 months. Research [6] evaluated the editing time of documentation using speech recognition (SR) to traditional typing (DT). Editing included the thorough examination, modification, or completion of clinical records in readiness for submission. Four studies shown a significant escalation in the duration of editing while using SR for generating reports [16, 18, 20, 21], whereas two studies indicated a little decrease. The changes in editing time varied from a significant increase of 1876.47% (17 seconds compared to 336 seconds using SR) to a modest decrease of 16.50% (303 seconds down to 253 seconds). Three researches investigated radiological records, while two studies focused on general clinical notes. Additionally, one study analyzed pathology reports. [20]

Eight researches were conducted to examine the turnaround time (TAT), which refers to the duration of the complete process from report generation to completion and submission [14, 19, 22, 27]. All of these studies consistently shown a reduction in TAT when using SR. The drop ranged from a high of 81.16% (1486 min down to 280 min) to a low of 16.41% (329 min down to 275 min). [19].

Precision of Speech Recognition

Multiple research have examined the precision of speech recognition (SR) as a method for entering data. These studies have focused on various factors, such as the number of mistakes per document for both SR and traditional data entry (DT), as well as the accuracy rate specifically for SR. Ten studies documented the accuracy rates of systematic reviews [14, 16, 18–20, 24, 25, 28–30]. Six researches conducted a comparison of the average mistakes per document between DT and SR. Specifically, studies [15, 16, 18, 20, 31, 32] all shown a significant rise in the number of errors while using SR. The mean mistakes per document using DT ranged from 0.02 to 0.40. The average mistakes per report generated using SR were much greater, ranging from 0.05 to 6.66. The average number of extra mistakes discovered per report when employing SR instead of DT varies from 0.03 to 19.53. [18]

Discussion

Speech recognition (SR) is a commonly used method of input for contemporary computer systems and has a long history of application in healthcare environments. Our research found that the data supporting the advantages and limits of using SR for clinical recording is limited, incomplete, and largely neutral towards its benefits. There is a lack of recent investigations that may benefit from more advanced SR technology.

In the late 1990s, three prominent companies introduced medical versions of their commercial speech recognition software. These systems were Dragon NaturallySpeaking with a medical add-on, IBM ViaVoice 98 with a general medicine vocabulary, and L&H Voice Xpress for medicine, General Medicine Edition. The early commercial distribution of SR in the 1990s generated a significant amount of excitement, which is seen in the perspectives expressed in the research and editorials from that time. The initial fervor for SR has gradually diminished, giving way to a more nuanced perspective that acknowledges both its advantages and limitations.[33-36]

The use of SR resulted in significant improvements in documentation speed at the system level, with a notable decrease in total turnaround time (TAT) for report preparation. The primary reason for this is the ability of SR-based systems to give information very instantaneously. This enhancement conceals the time burden of revising and creating documents, which is directly borne by the doctor. The successful implementation of technologies in clinical settings typically relies on the local expenses being compensated by local advantages. The limited adoption of speech recognition (SR) technology so far may be attributed, at least in part, to a discrepancy between the costs and benefits for clinicians who are responsible for generating reports.

These studies have indicated a slight improvement in the accuracy of SR over time. There are many technical factors contributing to this, such as advancements in microphone quality, speech recognition software packages, and computer hardware. Currently, several providers of speech recognition software claim accuracy rates as high as 99%. Despite achieving high accuracy rates, it is important to note that speech recognition (SR) may not be completely safe in a clinical setting. Various studies have identified a variety of mistakes, some of which have substantial clinical implications and might potentially damage patients. The reported errors encompassed a variety of issues, such as generating documentation for the incorrect patient, incorrect drug name or dosage, incorrect lab values, discrepancies in left/right anatomical information, medical inconsistencies, mismatches in age or gender, incorrect doctor name, incorrect date, usage of fabricated words and acronyms, irregular spacing, spelling mistakes, omissions, and duplications. DT had lower rates of clinically significant mistakes, largely attributed to trained and experienced transcriptionists who provided an extra safety check on the content of clinical documents.[37-39]

The experiments did not sufficiently capture the many factors that might possibly impact the operation of the SR system. The factors included in this context are: user training and expertise, speech recognition speed, microphone quality, author's accent or speech impairment, interruptions during dictation, degree of background noise, and other prevailing environmental circumstances. Without these data, it is necessary to use caution when making generalizations about the performance reported in these research and how it may translate to real-world situations. To clarify, achieving satisfactory results in controlled environments does not guarantee the same outcomes in real-world clinical situations.

These studies alone do not provide enough evidence to determine if SR is an efficient or successful method for creating clinical documentation. Additionally, it is unclear which situations and modes of usage are most suitable for SR. However, the extensive implementation of SR should provide proof of the existence of these advantages. Nevertheless, more study is required in several domains before this can be firmly confirmed.

Effect on clinical processes and results: None of the research included in this evaluation examined the influence of SR on clinical outcomes. The use of electronic documentation for the electronic health record is expected to primarily enhance organizational processes rather than clinical results. However, the utilization of 39 SR may lead to modifications in the amount and content of documents, which might potentially influence clinical choices. The implementation of SR will also lead to substantial changes in the business operations carried out by a company, and these modifications, along with their subsequent consequences, may need analysis.

The research included in this review found that although there were overall gains at the system level, there were also hidden costs for physicians. Although speech recognition (SR) may seem simpler than other conventional input methods, it seems to need more time when it comes to editing papers. If the use of SR further augments cognitive burden for physicians, it might potentially affect various clinical responsibilities, temporal efficiency, and rates of mistakes. [40]

The use of information technology is often linked to potential threats to patient safety. Technology not only introduces new categories of faults, but also presents new possibilities for user mistakes, which might differ depending on the particular technology, its users, their environment, training, and jobs. Considering the evident emergence of fresh mistakes linked to SR, every appraisal of advantages necessitates a careful consideration of potential damage.

This study did not include any research that compared speech recognition (SR) with the prevailing input method of keyboard and mouse. This may be due to the fact that dictation is widely used in fields like radiology, where most of the studies were done. In order to provide a conclusive evaluation of SR, it is necessary to compare it with other commonly used options such as keyboard input. The efficacy of SR in activities unrelated to documentation, including as order input, alert management, and patient handoffs, shows significant promise. Nevertheless, the bulk of the research failed to address these problems. Several of these activities may need the use of more limited restricted vocabularies, which might possibly simplify the work. However, there is a possibility that certain tasks may be more challenging due to the presence of more distractions or disturbances in the surroundings.

Various information technology platforms are used in clinical practice, ranging from conventional computer workstations to smart phones, tablet devices, and wearable devices like head-mounted displays and eyewear. These platforms may employ speech recognition (SR) technologies in diverse manners. The efficacy of SR is expected to differ across different systems.

Constraints of this review

Subgroup analyses were not possible due to the heterogeneity in research task, design, and population, as well as the small sample numbers. Although this feature enabled the examination of temporal patterns, it hinders the direct comparison of several researches. Likewise, the complexity of research designs improved with time, and earlier studies were often of inferior quality compared to more recent publications. Additional constraints include the extent, duration, and financial factors. There were few evaluations of the comparative efficiency of systematic reviews. When comparisons were conducted, they were performed using DT. The majority of research primarily concentrated on the production of radiological reports, with only a limited number investigating the generation of clinical data that would be included in the primary parts of a patient record. The use of speech for sophisticated purposes such as navigation and control was not investigated. Several potential factors that might impact real-world performance were not investigated nor documented.[41]

The research spanned over a period of more than twenty years, including many generations of SR technology. The SR systems used in this study were IBM MedSpeak and ViaVoice versions [20, 22, 24, 25, 28, 30]; Philips SP6000 and SpeechMagic versions [14, 19, 26]; Nuance's Dragon Naturally Speaking, Powerscribe, and SpeechMagic 2009+ versions [18, 21, 23, 32, 35]; AGFA's Talk Technology versions 16, 27, and 31; and AARK Systems, ASR Medispeak, the Dent Voice Prototype, and LTI. [15, 17, 29, 42]. The studies examined had limited and poorly reported cost-benefit assessments. Furthermore, due to major variances in factors such as clinical context, technology, installation, maintenance, and support expenses, most of these evaluations were not directly comparable.

Conclusion

This study demonstrates that SR has the potential to be a helpful tool for clinical recording. Nevertheless, it is essential to carefully consider the possible drawbacks of any improvements, such as the possibility of time penalties for physicians, the risk of additional mistakes, and the lack of clear cost-benefit analysis in some clinical environments. The available scientific data is quite limited, and several questions and untapped possibilities still exist. Although SR may not be feasible for every clinician in every situation, it is now unable to precisely define the specific activities and therapeutic environments in which its use is definitely advantageous and where it should potentially be abstained from. Smartphones and wearable devices, such head-mounted displays and eyewear, are well-suited for the therapeutic application of speech recognition (SR). Therefore, this topic deserves further study focus.

References

1. Johnson M Lapkin S Long V et al. . A systematic review of speech recognition technology in health care. BMC Med Inform Decis Mak. 2014; 14 (1): 94.
2. Herman SJ. Speech recognition and the creation of radiology reports. Appl Radiol.2004; 33 (5): 23 – 28.
3. Lawrence D. Can you hear me now? Voice recognition for the EMR has made big strides, and many say meaningful use requirements will accelerate adoption. Healthcare informatics: the business magazine for information and communication systems. 2009; 26(12).
4. Neustein A Markowitz JA. Mobile Speech and Advanced Natural Language Solutions. New York: Springer; 2013.
5. Bliss MF. Speech Recognition for the Health Professions: (using Dragon NaturallySpeaking). Upper Saddle River, N.J.: Prentice Hall; 2005.
6. Madisetti V. Video, Speech, and Audio Signal Processing and Associated Standards. Boca Raton, FL: CRC Press; 2009.
7. Paulett JM Langlotz CP. Improving language models for radiology speech recognition. J Biomed Inform. 2009; 42 (1): 53 – 58.
8. Gales M Young S. The application of hidden Markov models in speech recognition. Found Trends Signal Process. 2008; 1 (3): 195 – 304.
9. Eddy SR. What is a hidden Markov model?Nat Biotech. 2004; 22 (10): 1315 – 1316.
10. Indurkhya N Damerau FJ. Handbook of Natural Language Processing. Boca Raton, FL: CRC Press; 2012.
11. Lafferty J McCallum A Pereira FC. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. San Francisco, CA: Morgan Kaufmann Publishers Inc.; 2001.
12. Moher D Liberati A Tetzlaff J Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Int Med. 2009; 151 (4): 264 – 269.
13. Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull.1968; 70 (4): 213.
14. Hundt WS Scharnberg O Hold B et al. . Speech processing in radiology. Eur Radiol.1999; 9 (7): 1451 – 1456.
15. Rana DSH Shepstone G Pilling L Cockburn J Crawford M. Voice recognition for radiology reporting: is it good enough? Clin Radiol. 2005; 60 (11): 1205 – 1212.
16. Pezzullo JA Tung GA Rogg JM et al. Voice recognition dictation: Radiologist as transcriptionist. J Digital Imag. 2008; 21 (4): 384 – 389.

17. Mohr DNT Turner DW Pond GR et al. . Speech recognition as a transcription aid: a randomized comparison with standard transcription. JAMIA. 2003; 10 (1): 85 – 93.
18. Issenman RM, Jaffer IH. Use of voice recognition software in an outpatient pediatric specialty practice. Pediatrics. 2004; 114(3):e290-e293.
19. Vorbeck F Ba-Ssalamah A Kettenbach J Huebsch P. Report generation using digital speech recognition in radiology. Eur Radiol. 2000; 10 (12): 1976 – 1982.
20. Al-Aynati MM Chorneyko KA. Comparison of voice-automated transcription and human transcription in generating pathology reports. Arch Pathol Lab Med. 2003; 127 (6): 721 – 725.
21. Bhan SN Coblentz CL Norman GR Ali SH. Effect of voice recognition on radiologist reporting time. Can Assoc Radiol J. 2008; 59 (4): 203 – 209.
22. Rosenthal DIC Dupuy FS Kattapuram DE et al. Computers in radiology: computer-based speech recognition as a replacement for medical transcription. Am J Roentgenol. 1998; 170 (1): 23 – 25.
23. Chapman WW Aronsky D Fiszman M Haug PJ. Contribution of a speech recognition system to a computerized pneumonia guideline in the emergency department. Proc/AMIA Ann Symp AMIA Symp. 2000: 131 – 135.
24. Ramaswamy MR Chaljub G Esch O Fanning DD VanSonnenberg E. Continuous speech recognition in MR imaging reporting: advantages, disadvantages, and impact. Am J Roentgenol. 2000; 174 (3): 617 – 622.
25. Ilgner J Duwel P Westhofen M. Free-text data entry by speech recognition software and its impact on clinical routine. Ear, Nose Throat J. 2006; 85 (8): 523 – 527.
26. Koivikko MP Kauppinen T Ahovuo J. Improvement of report workflow and productivity using speech recognition—a follow-up study. J Digit Imaging. 2008; 21 (4): 378 – 382.
27. Krishnaraj A Lee JK Laws SA Crawford TJ. Voice recognition software: effect on radiology report turnaround time at an academic medical center. Am J Roentgenol. 2010; 195 (1): 194 – 197.
28. Kanal KM Hangiandreou NJ Sykes AM et al. Initial evaluation of a continuous speech recognition program for radiology. J Digit Imaging. 2001; 14 (1): 30 – 37.
29. Smith NT Brien RA Pettus DC Jones BR Quinn ML Sarnat A. Recognition accuracy with a voice-recognition system designed for anesthesia record keeping. J Clin Monitor.1990; 6 (4): 299 – 306.
30. Zemmel NJ Park SM Schweitzer J O'Keefe JS Laughon MM Edlich RF. Status of voicetype dictation for windows for the emergency physician. J Emerg Med.1996; 14 (4): 511 – 515.
31. McGurk S Brauer K Macfarlane TV Duncan KA. The effect of voice recognition software on comparative error rates in radiology reports. Brit J Radiol. 2008; 81 (970): 767 – 770.
32. Basma S Lord B Jacks LM Rizk M Scaranelo AM. Error rates in breast imaging reports: comparison of automatic speech recognition and dictation transcription. Am J Roentgenol. 2011; 197 (4): 923 – 927.
33. David GC Chand D Sankaranarayanan B. Error rates in physician dictation: quality assurance and medical record production. Int J Health Care Qual Assur. 2014; 27 (2): 99 – 110.

34. Quint LE Quint DJ Myles JD. Frequency and spectrum of errors in final radiology reports generated with automatic speech recognition technology. J Am Coll Radiol. 2008; 5 (12): 1196 – 1199.
35. Chang CA Strahan R Jolley D. Non-clinical errors using voice recognition dictation software for radiology reports: a retrospective audit. J Digit Imaging. 2011; 24 (4): 724 – 728.
36. Belton K Dick R. Voice-recognition technology: key to the computer-based patient record. J Am Med Record Assoc.1991; 62 (7): 27 – 32, 36, 38.
37. Clark S. Implementation of voice recognition technology at provenant health partners. J Am Health Inform Manag Assoc.1994; 65 (2): 34, 36, 38.
38. Nuance Communications. Dragon NaturallySpeaking 13 Premium Data Sheet - Nuance Communications. 2014: 2.
39. Coiera E. Guide to Health Informatics. 3rd edn. Boca Raton, FL: CRC Press; 2015.
40. Coiera E. Technology, cognition and error. BMJ Qual Saf2015; 24 (7): 417 – 422.
41. Coiera E Aarts J Kulikowski C. The dangerous decade. JAMIA. 2012; 19 (1): 2 – 5.
42. Chleborad KD Zvara T Hippmann K et al. Evaluation of voice-based data entry to an electronic health record system for dentistry. Biocybernetics Biomed Eng. 2013; 33 (4): 204 – 210.