The efficiency of voice recognition versus transcriptionist in Radiology

Amjad Sattar[1], Mahnoor Hafeez[2], Nida Rafiq[3], Ummey Aymen[3]

ABSTRACT

Objective: To estimate comparative efficiency of voice recognition system Voice Recognition System (VRS) and Medical Transcriptionist (MT) by qualitative and quantitative assessment of errors in Radiology Reports at Cross-Sectional Imaging.
Study Design: Prospective cross-sectional study
Place and Duration: Dow Institute of Radiology, DUHS; from 1st February to 30th April 2020.
Methodology Total consecutive criteria 201 cross-sectional reports were included in the study, these cases dictated by Radiologists and transcribed by Medical Transcriptionist were re- phonated on Voice Recognition System (VRS); the outcome of these reports were saved on Microsoft word files. Voice Recognition USB Headset and the microphone- both systems were analyzed for the functionality of VRS. Data was further categorized into CT and MRI long and short cases; error types and frequency were recorded.
Results: Mean Error rate (MER) of the Reports for voice recognition system (n=201) was 15.2% +/- 12.3 (S.D.) while that for Medical Transcriptionist (n=201) was 2% +/- 1.94 (S.D.). Independent t-test showed statistically significant greater Error Rate for VRS as compared to MT; [p-value 0.000]. A linear positive correlation was seen between no. of errors and total word count. There was no statistically significant difference between no. of errors for CT category as compared to MRI category, but regarding the error rate of MT and VRS Reports for MRI reports; there was a significant association of MER in the long cases as compared to short cases. In voice recognition system Reports, syntactic errors were found in a total of 184 Reports whereas semantic errors were found in a total of 82 reports. Typographical type error was the leading error seen in 175 Reports In Medical Transcriptionist reports; Syntactic and Semantic errors were seen in only a few Reports (14); [p-value- 0.022].
Conclusion: Medical Transcriptionist was found to have higher efficacy as compared to Voice recognition. VR has a significantly high frequency of error rate as compared to MT, deeming it unsuitable for implementation in cross-sectional imaging.
Keywords: Voice recognition software, Speech recognition software, Syntax, Semantics, Medical, Transcription.

How to Cite This:
Sattar A, Hafeez M, Rafiq N, Aymen U. The efficiency of voice recognition versus transcriptionist in Radiology. Isra Med J. 2021; 13(3): 192-196.

INTRODUCTION

The written report is the prime outcome of the Radiology Report and an error-free report is the prime requirement of the clinicians. Accurate Radiology Reporting is critical to health care quality and safety. Cross-sectional imaging takes the lead from other modalities, in the sense that its reports are more complex, lengthy, and time-consuming to compose. There are two speech modes for Radiology Reports: Voice recognition software (VRS) and Medical Transcriptionist (MT). Voice Recognition System (VRS) is a form of Artificial intelligence; it's a phenomenon of converting or transcribing acoustic human speech (i.e. sound waves) into a symbolic form of a human language such as English whereas MT is a medical language specialist, who deals in the process of transcription, and converts voice-recorded reports as dictated by physicians on Dictaphone (DP), into text format. DP is an electronic voice recorder analogous to the cell phone that saves and records voice files.

The introduction of VRS also known as automatic speech recognition (ASR) is a leading step and most economical step in the field of Radiology and a major technological advancement. The benefits of an efficient voice implementation system are twofold: it's a low budget plan as compared to the combined budget on the provision of Dictaphones and MT services. Meanwhile, it efficiently cuts report turnaround time (RTAT) and

1. Associate Professor of Radiology
2. Senior Instructor of Radiology
3. Assistant Professor of Radiology

Dow Institute of Radiology, DOW University of Health Sciences.

Correspondence:
Mahnoor Hafeez
Senior Instructor of Radiology, Dow Institute of Radiology, DOW University of Health Sciences.
Email: mahnoor.hafeez@yahoo.com

Received for Publication: March 20, 2021
1st Revision of Manuscript: June 07, 2021
2nd Revision of Manuscript: June 11, 2021
3rd Revision of Manuscript: August 11, 2021
4th Revision of Manuscript: August 29, 2021
Accepted for Publication: September 03, 2021

improves workflow[1,2]. Krishnaraj et al[3] in 2010 researched an academic setup, in which 30 faculty members were involved in the assessment of quantitative change in RTAT before and after VRS implementation and concluded that the average RTAT for the department before implementation of voice recognition was 28 hours. After the implementation of voice recognition, the average time was 12.7 hours. Having said this, VRS has hype for high error rates in reporting. VR dictation system is like a double-edged sword. Whilst there are many benefits, there are also many pitfalls[4]. Medical Transcriptionist (MT) Assembly is currently a basic component of the majority of Pakistani Radiology Departments. In one recent international study, among errors at the Speech recognition (SR), the medical transcriptionist-edited document (MT), and the physician's signed note stages, 15.8%, 26.9%, and 25.9%, respectively were clinically significant[5].

ASR is being increasingly used, but the quality has not been thoroughly studied in Asian countries. Error rate estimation is the best quality metric for testing the efficacy of the software. After successful implementation of the Voice Recognition System (VRS) in the General Radiography and Ultrasonography at our Institute, there was a plan to apply it in cross-sectional imaging as its usage significantly narrows report turnaround time (RTAT). The rationale was "Is VRS powerful enough to replace MT completely from our Imaging Institute". To assess the functionality and error rate of VRS software and its future implementation in cross-sectional imaging, the study was conducted at the tertiary care hospital. To the best of the authors' knowledge, a similar type of comparative analysis has not been conducted before in Asian Countries. So this study was conducted with an objective to estimate the comparative efficiency of voice recognition system (VRS) and Medical Transcriptionist (MT) by qualitative and quantitative assessment of errors in Radiology Reports at Cross-Sectional Imaging.

METHODOLOGY

This prospective cross-sectional study was conducted at Dow Institute of Radiology, DUHS; from 1st February to 30th April 2020, in which cases are selected from daily routine cross-sectional imaging via consecutive non-probability sampling. To avoid bias, all types of simple screening and complex contrast studies of all sections were included. There were no exclusion criteria. For the functionality of VRS, 'Dragon' Dictation software was installed by IT personnel and staff radiologists were trained on the personal mike to generate voice recording templates for accent recognition. The training session lasted for 30 minutes. Two Academic Radiologists, with 5 years of experience in Radiology, were involved in data collection. One used Voice Recognition USB Headset and the other used a VR microphone.

The cross-sectional reports transcribed by Medical Transcriptionist (MT) were re- phonated on VRS, the outcome of transcribed and re- phonated reports was saved on Microsoft word files. Screening CT, MR brain, and pyelogram cases comprised 'Short case exams'. 'Longcase exams' meant complex plain and contrast studies of all sections. Efficiency is also known as Efficacy is the ability to do things well, and is inversely proportional to 'No. of Error' and 'Error type' in each report, which was recorded. The error rate was defined as a percentage of words wrongly perceived by the software/ MT, per total word count in the report. The 'type' and 'frequency' of Error were also assessed. These errors included 'Syntax (grammatical)', 'Typographical (spelling)', and 'semantic errors'. Typographical error: also called typo error/ misprint – is mistyping due to human error. Syntax errors occur during the parsing of input code and are caused by grammatically incorrect statements. E.g. intense, the verb comes before a noun. Semantic errors occur during the execution of the code. It is the error in which words are perceived wrongly when the instructions are being executed; aka word substitution. E.g. Transcriptionist is perceived as 'descriptionist' (wrong spelling).

Data Analysis: The data were further categorized into CT and MRI cases, which were further recorded on an excel sheet. The numerical and categorical data was compiled in tabular and graphical form and analyzed on SPSS 20.0 software. An Independent t-test was applied to compare the means of the error rate of two groups. Chi-square test was applied to study the association between type of study, error type and rate in VRS and MT reports. Pearson's correlation was done to assess the relationship between the word count length of report and frequency of error.

RESULTS

There was total 201 cases comprising MRI (n=100, 49.7%) and CT (n= 101, 50.24%) were studies. A total of 7183 errors using VRS (100% of VRS reports) and 855 errors were identified in 201 MT reports (98% of MT reports). MRI studies included CT Short cases (n= 38, 18.9%), CT Long cases (n= 63, 31.3%), MRI Short cases (n= 72, 35.8%), and MRI Long cases (n= 28, 13.9%). For each category of Reports; the Mean Error rate (MER) of the Reports for VRS (n=201) was 15.2% +/- 12.3 (S.D.) and the Mean Error rate of the Reports for MT (n=201) was 2.00% +/- 1.94 (S.D.). 2 tailed independent t-tests showed a statistically significant difference between the error rate of two groups: VRS> MT; p-value 0.000, CI =95%. The mean word count of all Reports was 98 with an S.D. of +/-216. [Min. word count= 42; Max. word count= 570]. There was a positive correlation between the number of words in the Report and the number of errors in MT and VRS Reports; r=0.36, r=0.57 on XY scatter plot graph resp. (Figure-1). No statistical significance was found for the total no. of errors using two different devices of VR (p-value=0.001).

MER of CT (n= 101) and MRI (n=100) Reports for VRS were 16.34% and 14.08% respectively and for MT was 2.16% and 1.85% resp. There was no statistically significant difference between no. of errors for CT category as compared to MRI category [p value= 0.116]. As regard to CT category MER of MT and VRS Reports; there was no significant association of MER in the long cases; 1.85% and 14.08% resp.; as compared to short cases; 1.81% and 13.92 % resp.; [p value= 0.198]; But as regard to MER of MT and VRS Reports for MRI cases; there was a significant association of MER in the long cases; 2.25% and 17.06% resp.; as compared to short cases; 2.16% and 16.34%

resp.; [p value= 0.031]. These accounts are due to more lengthy Reports in long cases with more word count (max.343); Table-I

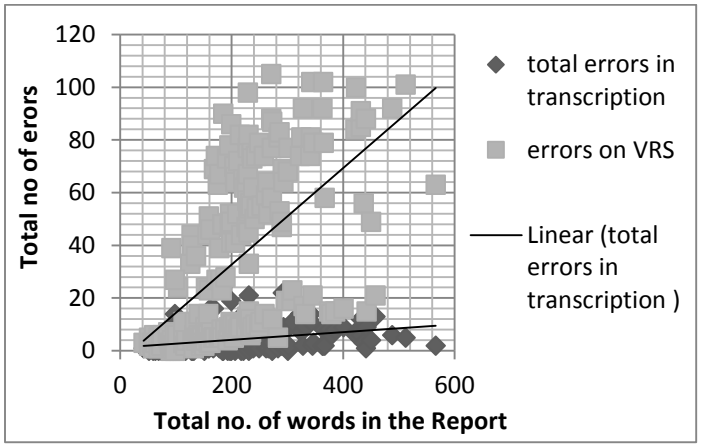

Figure-1: Graph shows the correlation between the total no. of errors and no. of words in the report.

Table-I: Comparative percentages for error rates in MT and VRS Reports with categorical distribution in short and long cases (n=201).

		Error Rate MT	Error Rate VRS	p value
Type of Study	MRI studies n= 100	(4/204) 2.16%	(37/204) 16.34%	0.116
	CT studies n= 101	(4/233) 1.85%	(35 /233) 14.08%	
CT	Short cases n= 38	1.81	13.92	0.198
	Long cases n= 63	1.85	14.08	
MRI	Short cases n= 72	2.16	16.34	0.031*
	Long cases n= 28	2.25	17.06	

*Chi-square test applied

*Denominator value in parenthesis implies mean of total word count in Reports

Table-II: Comparative frequencies of types of error for MT and VRS Reports (n=201)

Modality	MT		VRS		*p-value
	Frequency of Typographical error	Frequency of *'Other' error	Frequency of Syntactic error	Frequency of Semantic error	
MRI (100)	85/100	7/100	92/100	13/100	0.022
CT (101)	90/101	7/101	92/101	69/101	
Total	175/201 (87.1%)	14/201 (6.9%)	184/201 (91.5%)	82/201 (40.7%)	

*Chi-square test applied

*Other signifies total no. of semantic and syntactic error in Medical Transcriptionist generated Reports.

In VRS Reports, syntactic errors were found in 184/201 (91.5%) reports whereas semantic errors were found in 82/201 (40.7%) reports. In MT reports, typographical type error was the leading error seen in 175/201 (87.1%) Reports; syntactic and Semantic errors were seen in only a few Reports 14/201 (6.9%); only 4 Reports were error-free, whereas, in the VRS category, none of the reports was error-free. Significant association of typographical errors in MT generated Reports has been found as compared to Semantic and syntactic errors in the VRS generated Reports; [p-value- 0.022]; Table-II.

DISCUSSION

In this study, the authors tested commercially available software i.e. voice recognition versus medical transcription. The results of our study have shown a total of 7183 errors in 101 CT and 100 MRI reports using VRS (100% of VRS reports) and 855 errors were identified in 201 MT reports (98% of MT reports). Our results were a bit contrary to a recent local study of Hafeez et al, conducted in 2014[6] regarding retrospective analysis on South Asian non-native English speakers in 2014 in which a total of 50 errors were found in 1856 reports using VRS (3.37% of VRS reports) including 6 X-rays (19.35% of VRS errors), 11 Ultrasound [US] (35.45%), 6 Nuclear Medicine [NM] (19.35%), 8 Vascular and interventional Radiology [VIR] (25.8%); whereas, 19 errors identified in MT reports (2.03% of MT reports) including 3 X-rays (15.79% of DT errors), 6 US (31.58%), 4 NM (21.05%), 6 VIR (31.58%); though in both studies, error rate was high for VR as compared to MT.

McGurk at a British teaching hospital[7] conducted a study analogous to our study and concluded that VRS increases the number of errors in reports which are more likely to occur in noisy areas with a high workload and are more likely to be made by non-native radiologists.

In 2011, in Melbourne, the study was conducted[8] to estimate the error rates using VRS and 50 random finalized reports were scrutinized for errors including wrong word substitution, deletion, punctuation, other, and nonsense phrase. Reports were divided into two categories: computer radiography (CR = plain film) and non-CR (U/S, CT, MRI, nuclear medicine, and angiographic examinations). 11% of the reports in the CR group had errors. 2% of these reports contained non-sense phrases. 36% of the reports in the non-CR group had errors and out of these, 5% contained nonsense phrases.

Rosenthal et al[9] mentioned that the most frequent errors of MT personnel were misspellings, which do not occur with voice recognition; this fact was found in our study. Word substitution aka semantic error was the main error that occurs using the VRS, resulting from the built-in probabilities of its statistical language model. Word recognition errors are more frequent for users with foreign accents and with native English speakers. This type of error would make it difficult for anyone to edit the report other than the radiologist who dictated it. In our research, semantic errors (grammatically incorrect tense) were found in greater proportion as compared to syntactic error.

In one study of 2017, medical transcriptionists proofread 213,977 SRS-generated signed reports from 147 different radiologists. They concluded that change in the Dictation microphone did not affect the error rate, also noted in our study. Motyer et al in 2016 found 68 (75.56 %) 'Spelling and grammar errors', 20 (22.22 %) 'Missense' and 2 (2.22 %) 'Non sense' errors in their reports[10,11].

In 2015, Du Toit et al[12] compared the accuracy of SR and DT reports. 300 retrieved SR and the 300 DT reports were studied. Of the original 300 SR reports analyzed, 25.6% contained errors, with 9.6% being clinically significant. Only 9.3% of the DT reports contained errors, with 2.3% having a potential clinical impact. They concluded that Radiologists with second-language English were more likely to generate reports containing errors, but the level of seniority had no bearing.

Hodgson et al in 2015[13], reviewed literature assessing the impact of VR on dictation and editing time, document turnaround time (TAT), VR accuracy, error rates per document, and economic benefit in Twenty-three articles. They concluded that TAT consistently improved using VR compared to MT (16.41% to 82.34%); Document editing time increased using SR compared to DT in four of six studies; Mean error rate per report increased using SR (0.05 to 6.66) compared to MT (0.02 to 0.40), whereas in our research, it was 15.2% using VR compared to 2% using MT. In a systematic review over 19 years on 102 articles, the reported word error rates ranged from 7.4% to 38.7%[14]. Our study results are comparable to one recent study in which the transcription service was more accurate than the voice recognition program[15].

Hammana et al[16] reviewed the literature describing the impact of SR on report error rates and productivity in radiology departments from 1992 to 2013. They found that the percentage of reports containing at least one error varied from 4.8% to 89% for speech recognition, and from 2.1% to 22% for transcription.

This is the article highlighting issues of a diagnostic laboratory serving patients at the tertiary care level. To our knowledge, no such article has been published recently in Pakistan. It's an innovative study with a high novelty quotient. The technical study with the comparative outcome of machine versus human in the field of Radiology with qualitative and quantitative assessment of errors and usage of interesting speech lexicon was the strength of our study, though it has few limitations. The study was conducted by non-native English speakers as per the availability of the Faculty Members. It is obvious from the literature that error rates might be different for different VRS users. Second, relatively cheaper software was used in the study considering the available budget of the University. Its practical limitation is that it does not allow synchronized usage of VRS for dictation and mouse for scrolling CT and MRI images on DICOM viewers. In Pakistan, most institutes rely on MT and are adapted towards the usage of dicta-phones as compared to VRS, because of its user-friendly nature. The major issue dealing with VRS is the increase in the radiologists' speaking time as evident from the article of Pezzullo et al[17] that Reports dictated with voice recognition took 50% 'longer' to dictate despite being 24% shorter than those conventionally transcribed, there were 5.1 errors per case, and 90% of all voice recognition dictations contained errors before report signoff while 10% of transcribed reports contained errors and after sign-off, 35% of VR reports still had errors. In the international world, voice recognition has completely replaced Medical Transcriptionist; in our research, we found reversed results; the errors made by MT were due to no. of reasons including non-certification, dual jobs at a time, and lack of transcription training.

Observing the rising trends of this advanced technology especially in developing countries, the authors are quite hopeful that VRS is emerging as a powerful tool to completely replace MT from the market. Further research needs to be done in this regard, particularly regarding different kinds of software in developing countries. The typographical errors done by the Medical transcriptionist raised a call to the dept. to train and certify transcriptionists to increase their efficacy.

CONCLUSION

Medical Transcriptionist was found to have higher efficacy as compared to Voice recognition. Semantic and syntactic were the main errors found in the VR software.

AUTHOR'S CONTRIBUTION

Sattar A: Conceived idea, Designed research methodology, Final critical review of manuscript
Hafeez M: Data Analysis, Manuscript writing, Data compilation
Rafiq N: Data collection, Data compilation, Literature review
Aymen U: Data collection, Data compilation, Literature review

Disclaimer: The abstract of this research was presented as a research poster at the '36th Virtual Conference' of the Radiological Society of Pakistan.
Conflict of Interest: None.
Source of Funding: None.

REFERENCES

1. Yu D, Deng L. Automatic Speech Recognition. Springer London limited; 2016:1-2.
2. Sferrella SM. Success with voice recognition. Radiol Manage. 2003; 25(3):42-49.
3. 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.
4. Strahan RH, Schneider-Kolsky ME. Voice recognition versus transcriptionist: error rates and productivity in MRI reporting. J Med Imag Radiat Oncol. 2010; 54(5):411-414.
5. Zhou L, Blackley SV, Kowalski L, Doan R, Acker WW, Landman AB et al. Analysis of Errors in Dictated Clinical Documents Assisted by Speech Recognition Software and Professional Transcriptionists. JAMA Netw Open. 2018; 1(3):180530.
6. Hafeez M, Khalid D, Mirza W, Nadeem N. Radiology Reporting Errors; Voice Recognition Software versus Dictation Transcription Methods in Non-Native English Speakers. 19th PGME Conference - Enhancing Academic Standards for Quality Care. 2014 May Auditorium. Aga Khan University Hospital, Pakistan. Website: [https://www.researchgate.net/publication/269099421]. Accessed on: May 1, 2021.
7. McGurk S, Brauer K, Macfarlane TV, Duncan KA. The effect

of voice recognition software on comparative error rates in radiology reports. Br J Radiol. 2008; 81(970):767-770.
8. Chang CA, Strahan R, Jolley D. Non-clinical errors using voice recognition dictation software for radiology reports: a retrospective audit. J Digit Imag. 2011; 24(4):724-728.
9. Rosenthal DI, Chew FS, Dupuy DE, Kattapuram SV, Palmer WE, Yap RM, et al. Computer-based speech recognition as a replacement for medical transcription. Am J of Radiol. 1998; 170 (1):23-25.
10. Ringler MD, Goss BC, Bartholmai BJ. Syntactic and semantic errors in radiology reports associated with speech recognition software. Health Informatics J. 2017; 23(1):3-13.
11. Motyer RE, Liddy S, Torreggiani WC, Buckley O. Frequency and analysis of non-clinical errors made in radiology reports using the National Integrated Medical Imaging System voice recognition dictation software. Irish J Med Sci. 2016; 185 (4):921-927.
12. Du Toit J, Hattingh R, Pitcher R. The accuracy of radiology speech recognition reports in a multilingual South African teaching hospital. Biomed Cent Med Imag. 2015; 15(1):1-5.
13. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. J Am Med Inform Assoc. 2016; 23(e1):169-179.
14. Blackley SV, Huynh J, Wang L, Korach Z, Zhou L. Speech recognition for clinical documentation from 1990 to 2018: a systematic review. J Am Med Inform Assoc. 2019 1; 26 (4):324-338.
15. Zick RG, Olsen J. Voice recognition software versus a traditional transcription service for physician charting in the ED. Am J Emerg Med. 2001; 19(4):295-8.
16. Hammana I, Lepanto L, Poder T, Bellemare C, Ly MS. Speech recognition in the radiology department: a systematic review. Health Inf Manag. 2015; 44(2):4-10.
17. Pezzullo JA, Tung GA, Rogg JM, Davis LM, Brody JM, Mayo-Smith WW. Voice recognition dictation: radiologist as transcriptionist. J Digit Imag. 2008; 21(4):384-389.