**American Journal of Internal Medicine**
2019; 7(6): 147-150
http://www.sciencepublishinggroup.com/j/ajim
doi: 10.11648/j.ajim.20190706.12
ISSN: 2330-4316 (Print); ISSN: 2330-4324 (Online)

# The Adjunct of Voice Recognition to Medical Transcriptionist in Asian Countries–The Pros and Cons

**Amjad Sattar, Mahnoor Hafeez**

Dow Institute of Radiology, Dow University of Health Sciences, Karachi, Pakistan

**Email address:**
amj7777@hotmail.com (A. Sattar), mahnoor.hafeez@yahoo.com (M. Hafeez)

**To cite this article:**
Amjad Sattar, Mahnoor Hafeez. The Adjunct of Voice Recognition to Medical Transcriptionist in Asian Countries–The Pros and Cons. *American Journal of Internal Medicine.* Vol. 7, No. 6, 2019, pp. 147-150. doi: 10.11648/j.ajim.20190706.12

**Received**: July 14, 2019; **Accepted**: August 19, 2019; **Published**: October 26, 2019

**Abstract:** Voice recognition software (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 Dictaphone (DP) is an electronic voice recorder analogous to cell phone that saves and records voice files. The Radiologists believe that Report generation in Radiology is a daunting task, including reading scans, requiring analytical and observational skills, interpretation of findings, dictating cases, proof reading, re analyzing cases and signing off after corrections, especially when the case list is long. In solving this multi-step process, VRS and DP have emerged as handy tech savvy equipments for "automatic typing" of scans, with the involvement of Medical transcriptionist (MT) for timely generation of reports. In the past few decades, there has been considerable transition from manual hand signed reports to electronically generated reports. MT has been a closed companion of Radiologist, even in manually generated reports. There has been a threat to MT being replaced by VRS at tertiary care hospitals, because of its low economic impact. The pros and cons of tool are elaborated in this article with the survey of Radiology Institutes of Pakistan.

**Keywords:** Voice Recognition Software, VRS, Dictaphone, Medical Transcriptionist, MT, Clinical Practice

## 1. Introduction

A language is a set of valid sentences. We can break a sentence in two components: syntax and semantics. The term syntax refers to grammatical structure whereas the term semantics refers to the meaning of the vocabulary symbols arranged with that structure. Grammatical (syntactically valid) does not imply sensible (semantically valid). For example, the sentence "cows flow supremely" is grammatically ok (subject verb adverb) in English, but makes no sense. The grammatical mistake in the same would be "flow cows supremely"; the verb placed before the subject. The most recent citation from Studies in health technology and informatics in 2015 [1] highlighted syntactic and semantic errors due to VRS. Errors were classified as material if they were believed to alter interpretation of the report. "Immaterial" errors were sub classified as intrusion/omission or spelling errors.

In the last 5 years, Artificial intelligence (AI) techniques known as deep learning have delivered rapidly improving performance in image recognition, caption generation, and speech recognition. Radiology, in particular, is a prime candidate for early adoption of these techniques [2].

Dicta phone was first invented in 1923 whereas, Speech recognition becomes widely available in 1987, commercially (figure 1). VRS/ASR (Automated Speech Recognition) refers to the use of computer hardware and software-based techniques to identify and process human voice. The difficulty of the problem is affected by things such as (i) the requirement to transcribe words in spoken continuously rather than in isolation-e.g the words "six" and "seven" spoken in isolation are phonetically different to the phrase "six-seven" spoken continuously (ii) the ability to handle multiple speakers with different accents (iii) the application or not of training (iv) the use of low bandwidth speech (v) the requirement to perform the transcription in real time. The term "voice recognition" is often used to describe speech recogniton. Speech recognition aims to tell what someone is saying whereas voice recognition aims to tell who is saying it. The mechanism of voice recognition is shown in figure 2.

The rationale of this paper is either the adjunct of Voice Recognition to Medical Transcriptionist in Asian Countries is

a threat or boon? The objective of this paper is to make the reader aware of these new technologies and their utility in Asian countries in the field of Radiology.

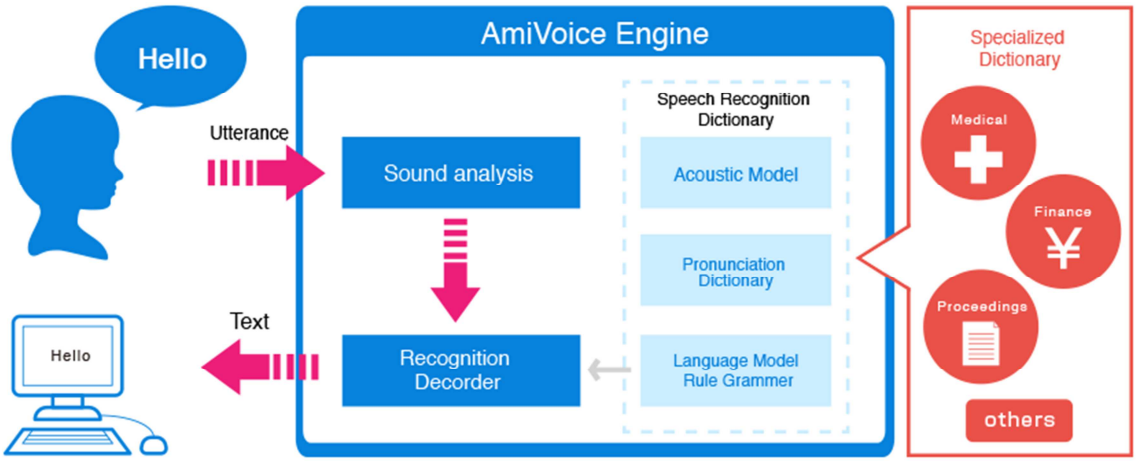


***Figure 1.*** *Flow chart showing mechanism of Voice recognition.*

## 2. Method

### *2.1. Workflow*

In narrating workflow at Dow University of Health Sciences (DHUS), in cross sectional imaging, there is an integrated complex, coordinated team of receptionist dealing with entry of patients, technologist for performing scans, IT personnel for dealing with PACS malware, transcriptionist and staff radiologists in the workflow - analogous to international setup. More than 15 radiologists are reporting them simultaneously on dicta phones which are transcribed by Team of 10 MT personnel at main campus. There is also a system of tele-radiography, scans dictated at peripheries of DUHS on computers, connected by local area network (LAN).

For the functionality of VRS in the institute, preliminary 'Dragon' Dictation software is installed by IT personnel and a short speech rehearsal of radiologist on personal mike to generate voice recording template for accent recognition by VRS. On the counterpart, there is a complex Dicta Phone - Medical Transcriptionist Assembly that is currently part and parcel of every Pakistani Radiology setup. MT is a medical language specialist, who deals in the process of transcription, and converts voice-recorded reports as dictated by physicians into text format. Then, after receiving the cassettes they transcribe them and send them to the Consultants in electronic/hard copy form.

### *2.2. City Survey*

Karachi being the metropolitan city of Pakistan has various major and minor radiology institutes in different areas. This data has been extracted from major institutes regarding utility of VRS and MT by authorize personnel.

***Figure 2.*** *Showing VRS mike and Dictaphone.*

***Table 1.*** *Utility of vrs and mt in karcahi pakistan (n=10).*

| S. No | Institute | MT | VRS | Reports Typed by Resident |
|---|---|---|---|---|
| 1 | Dow Institute of Radiology, DUHS | Y | Y | - |
| 2 | Civil Hospital and Trauma Centre | - | - | y |
| 3 | Agha Khan University Hospital | | y | - |
| 4 | PNS Shifa Hospital | Y | - | - |
| 5 | Jinnah Post graduate Medical center | - | - | Y |
| 6 | S.I.U.T | - | - | Y |
| 7 | Liaquat National Hospital | Y | - | - |
| 8 | Ziauddin University Hospital | Y | Y | - |
| 9 | Karachi X- rays | Y | - | - |
| 10 | Advanced Radiology Clinic | Y | - | - |

Y = Yes
- = No usage

# 3. Result

In this research, all mentioned are the Teaching Institutes of the city serving a huge load of community. Category 1-8 institutes are Tertiary Care Hospitals and last two are the radiology laboratories. The data in this study clearly shows that MT is taking the lead from VRS regarding its usage in 60% of the institutes (n=6), while VRS is used in in 30% of the institutes (n=3). Seldom, there has been a tradition of reports being typed by resident doctor in 30% of the institutes (n=3) because of cost effectiveness.

# 4. Discussion

Speech recognition software appeared in about 1952 but the first widely available example was the system developed by IBM. The following year, Dragon Systems released its first version of NaturallySpeaking for Microsoft Windows. There are three types: headset microphone, hand-held one and the desk microphone. Background noise and reverberation from hard surfaces can degrade the speech signal and cause multiple recognition errors, especially with an omnidirectional microphone [3]. During the past several years, the big share of health care wealth is invested on radiology services. This has forced the radiologist to deal with the task of providing increased services more efficiently in terms of both time and cost [4].

The radiologist, previously responsible only for recording a report onto tape and ensuring the integrity of the final report now becomes obligated to interact with the computer and to ensure the integrity of the transcription process as well as the accuracy of the final report [5].

The authors agreed to the statement of Sferrella SM that the benefits of efficient voice implementation system are twofold: it's a low budget plan as compared to combined budget on provision of dicta phones and medical transcriptionist services. Meanwhile, it efficiently cuts report turnaround time. [6]

On the positive aspect, VRS is beneficial for reducing radiology RTAT and improving workflow. Its importance was inferred from the study of Krishnaraj A [7] etal in 2010, in which 30 faculty members were ranked according to their RTAT before and after implementation of voice recognition and according to their percentage reduction in RTAT. The average RTAT for the department before implementation of voice recognition was 28 hours. After implementation of voice recognition, the average, it was 12.7 hours.

However, new technologies are not as smooth as they deemed. The citation in 2010 from Journal of Medical Imaging and Radiation Oncology [8] demonstrate that VRS is not an effective method of generating reports for MRI because of high error rates. In 2014, Hafeez [9] and his colleagues from Pakistan were convinced with the same fact and showed a retrospective analysis for the speakers who do not have English as a first language in a South Asian population. Total 50 errors were made in 1856 reports using VRS (3.37% of VRS reports) including 6 X-rays (19.35% of VRS errors), 11 US (35.45%), 6 NM (19.35%), 8 VIR (25.8%). Whereas, 19 errors identified in DT reports (2.03% of DT reports) including 3 X-rays (15.79% of DT errors), 6 US (31.58%), 4 NM (21.05%), 6 VIR (31.58%).

Another big issue dealing with VRS is the increase in the radiologists' speaking time as evident from the article of Pezzullo JA etal [10] 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 prior to report signoff while 10% of transcribed reports contained errors and after sign off, 35% of VR reports still had errors.

Rana DS [11] concluded that VRS is a viable reporting method for experienced users, with a quicker overall report production time despite an increase in the radiologists' time and a tendency to more errors for inexperienced users.

Bhan SN etal [12] concluded after research that for plain radiographs, radiologists took 13.4% more time to produce reports using VR, but for CT; there was no significant difference in reporting time identified between VR and CD (conventional dictation).

McGurk S conducted a study at British teaching hospital [13] regarding issue whether reports generated in a department of radiology contain more errors if generated using voice recognition (VR) software than if traditional dictation-transcription (DT) is used. Data collected included the type of report, site of dictation, the experience of the operator, and whether English was the first language of the operator. Total 1887 reports were reviewed. They 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, at Melbourne, Victoria, the study was conducted to ascertain the error rates of using a voice recognition (VR) dictation system [14]. 50 random finalized reports were scrutinized for errors in six categories namely, 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). Eleven percent of the reports in the CR group had errors. Two percent of these reports contained nonsense phrases. Thirty-six percent of the reports in the non-CR group had errors and out of these, 5% contained nonsense phrases. They concluded that VR dictation system is like a double-edged sword. Whilst there are many benefits, there are also many pitfalls.

In the clinical practice, different types of error made by VRS and by human transcription are observed as discussed by Rosenthal DI [15]. The most frequent errors of MT personnel are misspellings, which do not occur with voice recognition. Word substitution is the main error that occurs using the VRS, resulting from the built-in probabilities of its statistical language model. This type of error would make it difficult for anyone to edit the report other than the

radiologist who dictated it. Word recognition errors are more frequent for users with foreign accents and with native English speakers.

The major limitation regarding utility of VRS is non-synchronized simultaneous usage of VRS for dictation and mouse for manipulating images at current system of the institute. Initially VRS was used for cross-sectional imaging, but then the usage was abandoned because of failed speaking outcome while scrolling CT and MRI images on DICOM viewer. In comparison with practice at other imaging institutes of Pakistan, most institutes are adapted towards usage of dicta phones, because of its user friendly nature. The institute tackle with dual speaking system for different imaging modalities, DP for complex axial imaging, and VRS for relatively shorter reports including modalities of General Radiography, interventional radiology, fluoroscopy and Ultrasound.

# 5. Conclusion

The need of MT cannot be under estimated in multimodality busy setups. The discussion ended with the question that can a machine like VRS replace human transcriptionist? There has been the rising trend of VRS usage, negating the need of MT. From the recent trend in clinical practices during the last decades, it is very clear that voice recognition has a significant potential to fully replace the need of medical transcriptionist in the market. It's in no doubt a helping tool for rapidly expanding radiology. Having said this, in the current era of Pakistan analyzing the studies, VRS has to go through a lot of advancements with implementation of powerful software to secure its position in the market in future.

# References

[1] Ringler MD, Goss BC, Bartholmai BJ. Syntactic and Semantic Errors in Radiology Reports Associated With Speech Recognition Software. Stud Health Technol Inform. 2015; 216: 922.

[2] Tang A, Tam R, Cadrin-Chênevert A, Guest W, Chong J, Barfett J et al. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. Can Assoc Radiol J. 2018 May; 69 (2): 120-135.

[3] Graham-C Andrew. How I Use It: Voice recognition software. Occupational Medicine, Volume 62, Issue 4, 4 June 2012, Page 315.

[4] Mehta A. (2002) Voice Recognition. In: Dreyer K. J., Mehta A., Thrall J. H. (eds) PACS. Springer, New York, NY

[5] Gale B, Safriel Y, Lukban A, Kalowitz J, Fleischer J, Gordon D. Radiology report production times: voice recognition vs. transcription. Radiol Manage. 2001 Mar-Apr; 23 (2): 18-22.

[6] Sferrella SM. Success with voice recognition. Radiol Manage. 2003 May-Jun; 25 (3): 42-9.

[7] Krishnaraj A, Lee JK, Laws SA, Crawford TJ. Voice recognition software: effect on radiology report turnaround time at an academic medical center. AJR Am J Roentgenol. 2010 Jul; 195 (1): 194-7.

[8] Strahan RH, Schneider-Kolsky ME. Voice recognition versus transcriptionist: error rates and productivity in MRI reporting. J Med Imaging Radiat Oncol. 2010 Oct; 54 (5): 411-4.

[9] 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,* Aga Khan University Hospital, Pakistan, May 2014.

[10] Pezzullo JA, Tung GA, Rogg JM, Davis LM, Brody JM, Mayo-Smith WW. Voice recognition dictation: radiologist as transcriptionist. J Digit Imaging. 2008 Dec; 21 (4): 384-9.

[11] Rana DS, Hurst G, Shepstone L, Pilling J, Cockburn J, Crawford M. Voice recognition for radiology reporting: is it good enough? Clin Radiol. 2005. Nov; 60 (11): 1205-12.

[12] 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 Oct; 81 (970): 767-70.

[13] Rosenthal DI, Chew FS, Dupuy DE, Kattapuram SV, Palmer WE, Yap RM, Levine LA. Computer-based speech recognition as a replacement for medical transcription. AJR Am J Roentgenol. 1998 Jan; 170 (1): 23-5.

[14] Chang CA, Strahan R, Jolley D. Non-clinical errors using voice recognition dictation software for radiology reports: a retrospective audit. J Digit Imaging. 2011 Aug; 24 (4): 724-8. doi: 10.1007/s10278-010-9344-z.

[15] Bhan SN, Coblentz CL, Norman GR, Ali SH. Effect of voice recognition on radiologist reporting time. Can Assoc Radiol J. 2008 Oct; 59 (4): 203-9.