Article published online: 2021-07-30

Voice recognition is here comma like it or not period

Matthew A Fox, Carl J Aschkenasi, Arjun Kalyanpur
Department of Clinical Teleradiology, Teleradiology Solutions, Bangalore, India

Correspondence: Dr. Arjun Kalyanpur, Teleradiology Solutions Pvt Ltd., Plot # 7G, Opp Graphite India, Whitefield, Bangalore - 560 048, India. E-mail: arjun.kalyanpur@telradsol.com

Abstract

Voice recognition (VR) technology needs improvement, but is as integral to the current practice of radiology as Radiology Information Systems and Picture Archival and Communication Systems. In the 1990s, the radiology community gave VR technology a rather lukewarm reception, but since then it has emerged as the predominant method of radiology reporting in the United States. In this article, we examine how VR technology works, outline the positive and negative aspects of VR technology on work flow, identify common VR transcription errors and review the discussion on VR adoption in the recent literature. We add to the discussion our personal experiences in an international teleradiology group.

Key words: Acoustic; language; phoneme; transcription; voice recognition

Introduction

Over the past decade major advances in technology have not only impacted the way we image our patients, but also fundamentally changed the workflow in the practice of Radiology. With the emergence of Radiology Information Systems (RIS) and Picture Archival and Communication Systems (PACS), the process of reviewing radiology exams migrated from a film and paper based process to digitized data and media. A key component which has contributed to further streamlining and digitizing the radiology work flow has been the integration of Voice Recognition (VR) technology into the process of radiology report production. To better understand the role VR is playing in the evolving radiology workflow, we begin with a review of howVR technology works.

VR in a Nutshell

Voice recognition (VR) technology has made tremendous strides since its inception, attributed to IBM, which showcased a prototype of the technology at the 1962 Seattle World's Fair.[1] The "IBM shoebox" was a rudimentary computer that could perform speech recognition of 16 spoken words and the digits 0 through 9, producing either a printed output of simple arithmetic computations or serial activation of light bulbs. The basic elements of that prototype still serve modern-day VR systems. A microphone and sound card are used as an analog-to-digital-converter (ADC) of the continuous spoken sound waves that are transformed into vibrations which are then converted into digital data. The digital data is analyzed by the VR software, which breaks it down to the smallest sound segments used in speech, called phonemes. The VR software then runs the phoneme segmented digital data through a series of statistical algorithms in order to identify and contextualize strings of recorded phonemes into words and phrases found in its language library. In the last step, the VR software then outputs the digital data as a coherent text or generates a computer command [Figure 1].

Most VR software research and development is focused on the penultimate step in this process, phoneme contextualization. Two key components of this step are the acoustic model algorithms and the language model algorithm.[2] The acoustic model is composed of a library of word phoneme compositions, including common variations in pronunciation, which are used to statistically predict the most probable

Access this article online	
Quick Response Code:	Website: www.ijri.org
	DOI: 10.4103/0971-3026.120252

word spoken. The language model is composed of a library of domain-specific words and phrases used to statistically predict the most likely word sequence spoken. Generally, the acoustic model performs the first calculation of the expected word which is then checked for contextualization based on statistical probability calculated by the language model. This integrated model assures a higher degree of accuracy. For example, if the acoustic model predicts that the most likely words spoken are "No intracranial *pass* or hemorrhage," then the combined probabilities of both the acoustic and language model would be "No intracranial *mass* or hemorrhage.

Adoption of VR Technology in Radiology

A recent survey of over 1000 physicians from all fields of medicine revealed that over 75% use VR technology in their daily practice.[3] A similar informal poll in Diagnostic Imaging (Sept. 2011) yielded 80% of respondents as users, but 30% reporting dissatisfaction with the technology. Financial pressures and interests, most notably in the United States, have incentivized automated reporting rather than more costly human transcriptionists. Similar pressures have also meant a push for greater patient turnover in all clinical settings, especially in the emergency room (ER). VR has been promoted as a way to shorten the gap between scan time and report arrival. VR also dovetails with the digital migration of healthcare to electronic medical records (EMRs), and filmless technology supported by Radiology Information Systems (RIS) and Picture Archival and Communication Systems (PACS). Furthermore, the burgeoning practice of structured medical reporting also parallels the increased adoption of VR, where templates and macros can be used (for better or worse) to create modular reports for many commonly encountered diagnoses.[4] VR technology also provides an effective platform for promoting standardized reporting methods and terminology. In addition, the affordability of high-powered processors, as well as cheap memory, permits the use of VR technology with off-the-shelf computers and workstations.

But VR has always had its detractors, and still does, despite it being the heir apparent to traditional human transcriptionists. At the institution where one author (CJA) trained, there was a marked schism between those for and against the adoption of VR, and interestingly it was not predicated along the lines of older recalcitrant physicians versus their younger colleagues. Some doctors expressed a concern for the soon-to-be laid-off or "re-employed" transcriptionists. Others were doubtful about whether there would be a net gain in productivity when the chief activities of the transcriptionist were to be taken up by the radiologist. Many were concerned about errors, misappropriations, etc., that would be overlooked by a rushed radiologist, but could be caught in many instances by experienced transcriptionists.

Terror-Prune Retorting (Error-Prone Reporting)

The leading drawback for most radiologists using VR is that it can make many errors, some of which can be difficult to detect when the speaker himself proofs the report. Common voice recognition errors [Table 1] cause more than just frustration. These errors have potential medicolegal consequences and have real financial consequences as well. It has been estimated that the time spent correcting errors by VR-generated reports is worth about $80,000 dollars per radiologist, annually.[5] Reported VR transcription error rates vary from as low as 2% to as high as 90% prior to proofreading. For example, a British study in 2008 showed a statistically significant twofold difference in error rate between dictation-transcription (DT) and VR; however, the overall error rate was quite low at 3.8%.[6] This is vastly different from a report from Brown University which showed an error rate of 89% for VR compared to 10% for DT.[5] Interestingly, even after proofing by the radiologist, the VR error rate was still very high at 35%.

These numbers, however, are difficult to interpret as they reflect different VR software implemented and assessed in different ways along the radiology workflow. However, several types of errors are frequently encountered, including wrong-word substitution, nonsense phrases, and missing words.[7] Common examples of substitution of a wrong word were sound-alike words, such as "normally" substituted for "minimally," "ill-defined" substituted for "well-defined," or "location" substituted for "dislocation." Nonsense phrases may arise simply due to the VR software error in parsing of the phonemes. For example, the word "retroactive" would be incorrectly transcribed into "rich or active" or "valuable" into "buy you a bull." Another group

Table 1: Common voice recognition errors

Type	Example	Cause
Word substitution	Normally-Minimally	Similar sounding words
Nonsense phrases	Retroactive-Rich or active	Software parsing error
Missing words	There is (no) evidence of hemorrhage	Faulty microphone or hardware issue, fast speech
Number, date errors	23-33	Similar sounding words
Opposites	Hypodense-hyperdense	Similar sounding words

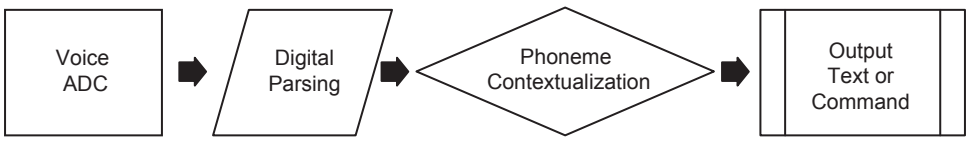

Figure 1: VR step model

of common errors involves numbers, measurements, and dates, which appear to be prone to VR transcription errors, and are difficult to catch when proofreading reports. One author (MAF) learned that a combination of improving the microphone hardware quality (i.e., USB versus audio jack), performing additional software VR training modules, as well as specific word training of the VR software led to an appreciable improvement in the VR transcription accuracy. Others have turned to an increased use of templates, reducing the words which are actually transcribed.

Impact on Training

Dr. Sanjeev Bhalla, chief of Thoracic Imaging and Assistant Program Director of the residency program at Mallinckrodt Institute of Radiology/Washington University in St Louis, raises further concerns regarding VR from a resident training standpoint.[8] Dr. Bhalla is concerned (and he is not alone) that VR diverts residents' attention from detailed review of the images, the nidus of their education, to a detailed review of the text in the dictation window. The time spent editing reports results in fewer cases covered and less training experience. He also observes that in a medical environment where the report is quickly available, direct consultations between radiologist and referring clinician decrease in both quantity and quality. When such consultations do occur, they are more apt to occur by phone, and revolve around the phrasing of a report rather than a personal exchange at the monitor in which imaging findings and clinical findings are compared.[9] The benefit of such direct discussion is better understanding of the clinical problem, its imaging manifestations, and ultimately a more meaningful final report reached through consensus impression. The learning potential of such interactions is enormous and may be jeopardized in a VR-mediated communication. Of course, among the counterarguments to these concerns would be the fact that as medical communicators, the report is our product, thus it is important to teach trainees to craft a cogent report with the tools that they will most likely use after they complete their training. A study conducted at Harvard Medical School concluded that in order for speech recognition to prove beneficial to radiology resident education, it must include adequate dedicated technical training and support including the use of specialized functions such as customizable templates/macros.[10] With improved user proficiency, the residents and attending physicians then need to be cognizant of properly incorporating the technology into the preview, review, and dictation process revolving image interpretation without losing sight of the necessary detailed visualization of the imaging findings by the residents.

The Need for Speed

Drawbacks notwithstanding, it is hard to do better than instantaneous, and that is chiefly the metric in which VR excels. Numerous studies have shown improved report turnaround time (TAT) with VR.[11,12] This is true for multiple languages and important particularly for our international practice, for English spoken by users whose first language is not English.[13] Rapid TAT is not only a clinically important issue in some areas, particularly in the ER environment, but also has become a critical component in cost containment in the ER, where rapid patient throughput is highly valued. In addition, continuity of care is significantly enhanced by the decreased report TAT, allowing for patient transfer with available imaging reports to level I trauma centers or other tertiary care facilities. Finally, it could further be surmised that rapid TATs may reduce risk in a litigious medicolegal environment.

Our Experience

The authors have worked with traditional human transcription systems, VR-based workflow, and a combination of these, where transcriptionists proofread VR reports. Furthermore, we were in training or early professional practice during the transition to widespread VR use. We have learned that radiologists must slow down and articulate words clearly to work effectively with VR despite "training" functions included in sophisticated VR software. Human transcriptionists can deal with rushed, run-on, or slurred dictation styles, which VR simply cannot. Although the software's adaptive technology learns from the user, this technology has its limitations. Even the most advanced user will find it necessary to compromise or work around some small inherent foible of the software in order to maximize its advantages. In addition, we have learned to appreciate the use of templates in reporting, a convenience which saves time during routine cases, enabling us to devote more attention to complex cases requiring carefully considered language. In order to leverage the advantages of VR over its disadvantages [Table 2] we have implemented a two-tiered approach in our practice with VR as the "front-end" reporting modality, followed by human editing. The medical transcriptionist still adds value in that he can clean up errors missed by both report author and VR software and ensure that exams are properly identified with the relevant patients and that specific formatting requests of a client are met.

Table 2: Pros and cons of using voice recognition

Positive	Negative
Decreased report turnaround time	More transcription errors
Decreased cost over time	Training period for software adaptation
Integrated into RIS/EMR for smooth workflow	Sensitivity to local accents
Facilitates standardized reporting nomenclature	Lost productivity of radiologist for proofreading
Improved patient throughput	Distraction from resident education
Improved continuity of care for transferred patients	Curtails consultations between radiologist and referring clinician

RIS/EMR: Radiology information systems/Electronic medical records

For the demands of our practice, which covers numerous modalities for globally dispersed clients with different needs, expectations, and medical cultures, and globally dispersed radiologists with differing approaches to practice and differing accents, we have found VR an indispensable tool (but not without its challenges), which helps unify our group with a common reporting interface.

References

1. IBM.com. Available from: http://www03.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html [Last accessed on 2011 Nov 16].
2. Schwartz LH, Kijewski P, Hertogen H, Roossin PS, Castellino RA. Voice recognition in radiology reporting. AJR Am J Roentgenol 1997;169:27-9.
3. Torrieri, M. Talk vs Type: Taking another look at voice recognition. Physician Practice. vol. 21. No. 7. Available from: http://www.physicianspractice.com/voicerecognitiondictation/content/article/1462168/1889679 [Last accessed on 2011 Jul 8].
4. Liu D, Zucherman M, Tulloss WB Jr. Six characteristics of effective structured reporting and the inevitable integration with speech recognition. J Digit Imaging 2006;19:98-104.
5. Pezzullo JA, Tung GA, Rogg JM, Davis LM, Brody JM, Mayo-Smith WW. Voice recognition dictation: Radiologist as transcriptionist. J Digit Imaging 2008;21:384-9.
6. 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;81:767-70.
7. 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;5:1196-9.
8. Personal communication between Dr. Sanjeev Bhalla with author (CJA) on 10.11.2011
9. Hayt DB, Alexander S. The pros and cons of implementing PACS and speech recognition systems. J Digit Imaging 2001;14:149-57.
10. Gutierrez AJ, Mullins ME, Novelline RA. Impact of PACS and voice-recognition reporting on the education of radiology residents. J Digit Imaging 2005;18:100-8.
11. Kauppinen T, Koivikko MP, Ahovuo J. Improvement of report workflow and productivity using speech recognition--a follow-up study. J Digit Imaging 2008;21:378-82.
12. Krishnaraj A, Lee JK, Laws SA, Crawford TJ. Voice recognition software: Effect on radiology report turn around time at an academic medical center. AJR Am J Roentgenol 2010;195:194-7.
13. Akhtar W, Ali A, Mirza K. Impact of a voice recognition system on radiology report turnaround time: Experience from a non-english-speaking South Asian Country. AJR Am J Roentgenol 2011;196:W485.

Cite this article as: Fox MA, Aschkenasi CJ, Kalyanpur A. Voice recognition is here comma like it or not period. Indian J Radiol Imaging 2013;23:191-4.
Source of Support: Nil, **Conflict of Interest:** None declared.