



Incidence of speech recognition errors in the emergency department



Foster R. Goss^{a,b,*}, Li Zhou^{c,d,e,g}, Scott G. Weiner^f

^a University of Colorado, Department of Emergency Medicine, Aurora, CO, United States

^b Tufts Medical Center, Department of Emergency Medicine and Clinical Decision Making, Boston, MA, United States

^c Division of General Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA, United States

^d Clinical & Quality Analysis, Partners HealthCare System, Boston, MA, United States

^e Clinical Informatics, Partners eCare, Partners HealthCare System, Boston, MA, United States

^f Division of Health Policy Translation, Department of Emergency Medicine, Brigham and Women's Hospital, Boston, MA, United States

^g Harvard Medical School, Boston, MA, United States

ARTICLE INFO

Article history:

Received 1 May 2015

Received in revised form 15 April 2016

Accepted 24 May 2016

Keywords:

Speech recognition
Emergency medicine
Patient safety

ABSTRACT

Background: Physician use of computerized speech recognition (SR) technology has risen in recent years due to its ease of use and efficiency at the point of care. However, error rates between 10 and 23% have been observed, raising concern about the number of errors being entered into the permanent medical record, their impact on quality of care and medical liability that may arise. Our aim was to determine the incidence and types of SR errors introduced by this technology in the emergency department (ED).

Setting: Level 1 emergency department with 42,000 visits/year in a tertiary academic teaching hospital.

Methods: A random sample of 100 notes dictated by attending emergency physicians (EPs) using SR software was collected from the ED electronic health record between January and June 2012. Two board-certified EPs annotated the notes and conducted error analysis independently. An existing classification schema was adopted to classify errors into eight error types. Critical errors deemed to potentially impact patient care were identified.

Results: There were 128 errors in total or 1.3 errors per note, and 14.8% (n = 19) errors were judged to be critical. 71% of notes contained errors, and 15% contained one or more critical errors. Annunciation errors were the highest at 53.9% (n = 69), followed by deletions at 18.0% (n = 23) and added words at 11.7% (n = 15). Nonsense errors, homonyms and spelling errors were present in 10.9% (n = 14), 4.7% (n = 6), and 0.8% (n = 1) of notes, respectively. There were no suffix or dictionary errors. Inter-annotator agreement was 97.8%.

Conclusions: This is the first estimate at classifying speech recognition errors in dictated emergency department notes. Speech recognition errors occur commonly with annunciation errors being the most frequent. Error rates were comparable if not lower than previous studies. 15% of errors were deemed critical, potentially leading to miscommunication that could affect patient care.

© 2016 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

1.1. Background

Physician use of computerized speech recognition (SR) technology has risen in recent years due to its ease of use and efficiency at the point of care. Nearly half of all licensed U.S. physicians use SR to enter information into the electronic health record (EHR) in their practice via a variety of methods [1]. Traditionally, a voice

dictation generated by the physician is sent to a medical transcriptionist who manually transcribes the document and sends it back to the physician for review. Front-end SR (or SR-generated documentation) occurs when a physician dictates into a text field in the EHR or text document using SR software and edits the dictation in real-time before saving it. Back-end SR (or SR-assisted transcription) occurs when a physician dictates and the recorded transcription is automatically processed by SR software that sends it to a human transcriptionist to review and finally to the physician for review. While front-end SR is the most likely method used in emergency department (ED) and what was used in this study, other methods do exist and often depend on the type of SR systems supported within their respective institution.

* Corresponding author at: Department of Emergency Medicine, University of Colorado, United States.

E-mail address: Foster.Goss@ucdenver.edu (F.R. Goss).

Table 1
SR Error Types, Descriptions and Examples.

Error type	Description	Example ^a
Annunciation errors	Occurring due to speaker mispronunciation	<i>He will see her on reactive</i> Pupils equal round and reactive
Dictionary errors	Resulting from missing terms	<i>solo bricks</i> Celebrex
Suffix errors	Caused by misrecognition of appropriate tenses of a word	<i>markedly</i> marked
Added words	Word added	8 year old male <i>the</i> history
Deleted words	Word deleted	O2 saturation....percent
Homonym errors	Resulting from substitution of a phonetically identical word	Nares <i>or</i> two days for two days
Spelling errors	Occurred only in human- transcribed notes and not with speech- recognized notes	and get her sugars yet
Nonsense errors	Resulting from words or phrases whose meaning could not be understood by examining the context	<i>Patient up been admitted for stable gait</i>
Critical errors	Were deemed to be “critical” if they could potentially impact patient care	pulse 175, respiration 32, <i>temperature 12.9</i> , room air O2 saturation percent.

^a Actual examples of SR errors identified within this study with the exception of Dictionary and Suffix errors, of which our sample had none.

1.2. Errors rates using speech recognition

Despite the advantages of SR technology, high error rates ranging from 10 to 23% have been observed in clinical documents generated by this technology [2], raising concern about the number of errors being entered into the permanent medical record, their impact on quality of care and the medical liability that may arise. To date, there have been few studies published on the use SR in ED [3–5]. A recent study by Zick et al. evaluated the accuracy and cost savings of traditional voice dictation as compared to a real-time SR software and observed high accuracies of 99.7% and 98.5% respectively [5]. Turnaround time was faster using the SR software as compared to traditional transcription and SR generated notes were less costly. While accuracy was reported, the types of errors that occurred were not systematically classified. In this pilot study, we sought to systematically classify and identify the incidence of SR errors in ED using a predefined classification schema by Zafar et al [6]. To the best of our knowledge, this has never been studied before in the ED. This work attempts to add to a much wider discussion on the use of technology its impact on patient care and safety.

1.3. Case report

A 25-year old female presented to the emergency department with an abscess on her arm. On questioning, the patient mentioned that she had missed her period. The patient was evaluated by the attending physician and a note was dictated. The physician commented in their note that the patient had missed her period. The software interpreted the physician’s reference to “period” as a punctuation mark “.”. She returned the following day with a worsening cellulitis on her arm and a colleague of the first doctor prescribed an antibiotic that was contraindicated during pregnancy.

2. Methods

2.1. Data collection and sampling

This study was conducted in an urban academic emergency department located in Boston with 42,000 patient visits per year. A random sample of 360 attending emergency physician (EP) notes recorded from January to June 2012 (60 notes/month) was collected from the ED EHR system. Notes could be either hand-typed or dictated using Dragon Medical Software 10.0 or 10.1 (Nuance Communications Inc.). Dictations were performed using the Nuance PowerMic II and primarily dictated in the Emergency Depart-

ment, an environment with higher ambient noise than the office setting. Only dictated notes were analyzed. To ensure each sample contained a representative number of words to analyze, we excluded sentences with less than 50 words (e.g. fragments or partial/incomplete dictations). Sample size was calculated using pilot data [7] and the software PASS (Power Analysis and Sample Size Software, version 11) [8]. We determined a sample size of 100 notes yielded an acceptable 95% confidence interval for notes containing critical errors of 14.3–31.4%, respectively. IRB approval was obtained for this study and determined it to be exempt as it posed no more than minimal risk to patient and all information was de-identified.

2.2. Error analysis

Error analysis was conducted independently by the two reviewers. Notes were reviewed and annotated using Knowtator [9], a text annotation tool built upon Protégé [10], an open-source ontology editor from Stanford University. We created a classification schema in Protégé based on Zafar et al. [6] (Table 1). Errors were deemed to be “critical” if they were believed by the reviewing physicians to potentially impact patient care. Identified errors were then reviewed and then jointly classified by the two reviewers and inter-annotator agreement calculated using Knowtator. Summary statistics were generated.

3. Results

Two board-certified EPs reviewed the notes and excluded those that were not dictated (n = 55), partially dictated (n = 7) or less than 50 words (n = 198). In total, 100 notes were included, dictated by 12 providers with a mean of 8.3 (SD 4.3) notes per provider. The number of words in the notes ranged from 50 to 500 with the mean being 140.0 (SD 74.9). Inter-annotator agreement on the jointly classified errors was 97.8%. Overall, 71% of the notes contained errors. There were 128 errors in total or 1.3 errors per note. Annunciation errors were the highest (53.9%) followed by deletions (18.0%), added words (11.7%), and nonsense errors (10.9%). Homonyms and spelling errors were lower at 4.7% and 0.8%, respectively. There were no suffix or dictionary errors. Of these errors, 14.8% were judged to be critical errors. 15% of notes contained one or more critical errors. A summary of these types of errors are shown in Table 2 and examples of critical errors in Table 3.

Download English Version:

<https://daneshyari.com/en/article/516057>

Download Persian Version:

<https://daneshyari.com/article/516057>

[Daneshyari.com](https://daneshyari.com)