Title: A systematic review of speech recognition technology in health care.

Running head: Speech recognition technology in health care.

Authors:

Maree JOHNSON, PhD (Epidemiology & Population Health) (ANU).
Email: maree.johnson@acu.edu.au

Samuel LAPKIN, RN, PhD, BN (Hons).
Email: samuel.lapkin@cqu.edu.au

Vanessa LONG, PhD.
Email: vanessawlong@gmail.com

Paula SANCHEZ, BNursing, GradDip Nursing.
Email: p.sanchez@uws.edu.au

Hanna SUOMINEN, M.Sc. (Applied Mathematics, incl. B.Sc. (Mathematics), University of Turku, Finland), PhD (Computer Science, University of Turku, Finland).
Email: Hanna.Suominen@nicta.com.au

Jim BASILAKIS, MB.BS.BSc(Med).MBiomedE. GradDipCS.
Email: j.basilakis@uws.edu.au

Linda DAWSON, BSc (Syd), MSc.(UTas), PhD (Information Systems, Monash University).
Email: lindad@uow.edu.au

Corresponding Author: Professor Maree Johnson, Associate Dean (Research), Faculty of Health Sciences, Australian Catholic University, 40 Edward Street, North Sydney, NSW 2060. Australia. Email: maree.johnson@acu.edu.au.
Abstract

Background

Objectives: To undertake a systematic review of existing literature relating to speech recognition technology and its application within health care.

Method

Study design: A systematic review of existing literature from 2000 was undertaken.

Inclusion criteria were: all papers that referred to speech recognition (SR) in health care settings, used by health professionals (allied health, medicine, nursing, technical or support staff), with an evaluation or patient or staff outcomes. Experimental and non-experimental designs were considered.

Data sources: Six databases (Ebscohost including CINAHL, EMBASE, MEDLINE including the Cochrane Database of Systematic Reviews, OVID Technologies, PreMEDLINE, PsycINFO) were searched by a qualified health librarian trained in systematic review searches initially capturing 1,730 references. Fourteen studies met the inclusion criteria and were retained.

Results

The heterogeneity of the studies made comparative analysis and synthesis of the data challenging resulting in a narrative presentation of the results. SR, although not as accurate as human transcription, does deliver reduced turnaround times for reporting and cost-effective reporting, although equivocal evidence of improved workflow processes.

Conclusions

SR systems have substantial benefits and should be considered in light of the cost and selection of the SR system, training requirements, length of the transcription task, potential use of macros and templates, the presence of accented voices or experienced and in-experienced typists, and workflow patterns.
**Keywords** Nursing, systematic review, Speech Recognition, Interactive Voice Response Systems, Human Transcriptions, health professionals.
Background

Introduction

Technologies focusing on the generation, presentation and application of clinical information in healthcare, referred to health informatics or eHealth solutions [1, 2] have experienced substantial growth over the past 40 years. Pioneering studies relating to technologies for producing and using written or spoken text, known as computational linguistics, natural language processing, human language technologies, or text mining, were published in the 1970s and 1980s [3-10]. Highlights of 1990s and early 2000s include the MedLEE Medical Language Extraction and Encoding System to parse patient records and map them to a coded medical ontology [11] and the Autocoder system to generate medical diagnosis codes from a patient record [12]. Today, a literature search using Pubmed for computational linguistics, natural language processing, human language technologies, or text mining recovers over 20,000 references.

Health informatics or eHealth solutions enable clinical data to become potentially accessible through computer networks for the purposes of improving health outcomes for patients and creating efficiencies for health professionals [13-16]. Language technologies hold the potential for making information easier to understand and access [17].

Speech recognition, in particular, presents some interesting applications. Speech recognition (SR) systems compose of microphones (converting sound into electrical signals), sound cards (that digitalise the electrical signals) and speech engine software (that convert the data into text words) [18]. As early as 1975 speech recognition systems were described ‘in which isolated words, spoken by a designed talker, are recognized through calculation of a minimum prediction residual’ [19] reporting a 97.3 per cent recognition rate for a male speaker. Applications have been demonstrated in radiology [20] with the authors noting a reduction in turnaround time of reports from 15.7 hours to 4.7 hours, although some
difficulties with integration of systems have also been identified [21]. Document processing within endocrinology and psychiatry including physicians and their secretaries also demonstrated improvements in productivity [22]. Similar approaches have recently been applied in the reporting of surgical pathology with improvements in ‘turnaround time from 4 to 3 days’ and ‘..cases signed out in 1 day improved from 22% to 37%’ [23]. These authors also alluded to the issue of correction of errors and the use of templates [23] for processing of information.

Although systematic reviews of health informatics [24-27] have been conducted, surprisingly we were unable to locate such a review on speech recognition in health care.

**Aim**

The aim of this study was to undertake a systematic review of existing literature relating to SR applications, including the identification of the range of systems, implementation or training requirements, accuracy of information transfer, patient outcomes, and staff considerations. This review will inform all health professionals about the possible opportunities and challenges this technology offers.

**Methods**

All discoverable studies published in the refereed literature from the year 2000 and in English language only were included in the review. We believed that only studies from 2000 onwards would use speech recognition technology that was sufficiently accurate to be suitable for health care settings. Papers were included if they referred to speech recognition in health care settings, being used by health professionals (allied health, medicine, nursing, technical or support staff), with an evaluation or patient or staff outcomes. All research designs, experimental and non-experimental, were included. Studies were excluded if they were opinion papers or describing technical aspects of a system without evaluation. Methods
for searching the literature, inclusion criteria, and general appraisal and analysis approaches were specified in advance in an unregistered review protocol.

Data sources (Search strategy)

Six databases (Ebscohost including CINAHL, EMBASE, MEDLINE including the Cochrane Database of Systematic Reviews, OVID Technologies, PreMED-LINE, PsycINFO) were searched by a qualified health librarian trained in systematic review searches, using the following search terms: "automatic speech recognition", "Speech Recognition Software", "interactive voice response systems", "((voice or speech) adj (recogni* or respon*)).tw.", "(qualitative* or quantitative* or mixed method* or descriptive* or research*).tw.". It should be noted that EMBASE includes 1000 conference proceedings (grey material) also. In addition, a search was undertaken for grey literature in Open Grey. Examples of the searches undertaken from three major databases are presented in Table 1.

Insert table 1

Results

Selection of studies

The search identified 1,730 references to publications that were published in or after 2000. There were 639 duplicates in these 1,730 references which were removed resulting in 1,091. Some 1,073 papers were not found to be relevant as they reflected other topics or applications such as: auditory research (65), cochlear implant or hearing instrument (174), conversations, or multiple speakers (12), discrete speech utterance (2), impaired voice (150), informal research notes including comments or response (6), interactive voice response (199), speech perception (53), synthesized speech (4), thesis (1), and other irrelevant topics (340). The remaining 18 were examined using the inclusion criteria by two independent reviewers and 14 papers (see Figure 1) were retained. All identified abstracts were reviewed by two reviewers, and a third where there was disagreement. The relevant full text of the article was
obtained and then if the paper met the eligibility criteria (checked by two reviewers) the study was included. Inclusion criteria were: referred to speech recognition in health care settings, used by health professionals (allied health, medicine, nursing, technical or support staff), with evaluation of patient or staff outcomes.

Insert Figure 1

The quality of each eligible study was rated by two independent reviewers using the Mixed Methods Appraisal Tool (including a range of quantitative designs the focus in this review) [28]. The scores for the included studies ranged from 4 to 6 out of a possible maximum of 6 [22, 29] (See Table 2). Data were extracted from the relevant papers using a specifically designed data extraction tool and due to the nature of the content reviewed by two reviewers.

Insert Table 2
Description and methodological quality of included studies

Of the fourteen studies retrieved, one was a randomised controlled trial (RCT) [22]; ten were comparative experimental studies [18, 20, 23, 29-35] and most of the remaining were descriptive studies predominately using a survey design [36-38].

The studies were conducted in hospitals or other clinical settings including: emergency departments [29, 35], endocrinology [22]; mental health [22, 30], pathology [18, 23], radiology [20, 33, 34, 38]; and dentistry [32]. However, one study was carried out in a laboratory setting simulating an operating room [36].

The health professionals or support staff involved were: nurses [29], pathologists [23], physicians [22, 29, 30, 35, 37], radiologists [18, 33, 38], secretaries [22], transcriptionists [18, 22] and undergraduate dental students [32]. In one study no participants were identified [36].

Training varied between studies with some studies providing data based on minimal training 5 minutes [29] to 30 minutes [23] to 6 hours [22]. One study emphasised the need for one to two months use before staff were familiar with SR [30].

The majority of the papers focused on systems that supported English language, however other languages such as Finnish [33] and Danish [36] were also investigated. Participants in two studies were non native English speakers although they transcribed documents into English [18, 38].

The quality scores for the studies ranged from two studies at 4 [29, 36], six studies at 5 [18, 30, 32, 34, 35, 38], and six studies at 6 [20, 22, 23, 31, 33, 37], with 6 being the maximum score possible (see Table 2).

Outcomes of the studies

The main outcome measures in the included studies were: productivity including report turnaround time [20, 22, 23, 29, 33-35]; and accuracy [18, 22, 29, 35].
The findings of the included studies were heterogeneous in nature, with diverse outcome measures, which resulted in a narrative presentation of the studies (See Table 3).

Results

Productivity

The search strategy yielded six studies that evaluated the effect of SR systems on productivity—report turnaround time (RTT), or proportions of documents completed within a specified time period. Overall, most papers [22, 29, 33-35] reported significant improvement in RTT with SR. Two studies reported a significant reduction of RTT when SR was used to generate patient notes in an emergency department (ED) setting [29] and clinical notes in endocrinology [22]. A longitudinal study (20,000 radiology examinations) indicated that using SR reduced RTTs by 81% with reports available within one hour increasing from 26% to 58% [33]. Similarly, the average RTT of surgical pathology reports was reduced from four days to three days with increases in the proportion of reports completed within one day (22% to 36%) [23]. Zick and Olsen reported the reduction in RTT achieved by using SR in ED resulted in annual savings of approximately $334,000 [35].

Insert table 3

Results of another study reported significant differences in RTT between SR systems produced by different companies. The authors reported that Dragon software took the shortest time (12.2 mins) to dictate a 938-word discharge report followed by IBM and L & H [31].

Quality of reports

The quality of the reports in seven studies was determined by comparing errors or accuracy rates [18, 23, 29, 31, 35, 36, 38]. Taken together results from these studies suggest that human transcription is slightly more accurate than SR. The highest reported average accuracy rate across the included studies was 99.6% for human transcription [18] compared
to 98.5% for SR [35]. However, an ED study found that reports generated by SR did not have grammatical errors while typed reports contained spelling and punctuation mistakes [29].

Evidence from the included studies also suggests that error rates are dependent on the type of SR system. A comparison of three SR systems indicated that IBM ViaVoice 98 General Medical Vocabulary had the lowest overall error rates compared with Dragon Naturally Speaking Medical Suite and L&H Voice X-press for Medicine, General Medicine Edition, when used for generating medical record entries [31]. A similar comparative analysis of four dental SR applications reported variation with regards to: time required to complete training, error rates, total number of commands required to complete specific tasks, dental specific functionality, and user satisfaction [32].

System design

Some SR systems incorporated generic templates and dictation macros that included sections for specific assessment information such as chief complaint, history of present illness, past medical history, medications, allergies and physical examination [22, 35]. Other researchers used SR systems with supplementary accessories for managing text information such as generic templates [22], medical or pathology terminology dictionary [18, 20, 31, 35] Radiology Information System (RIS) [34] and Picture Archiving and Communication System (PACS) [33]. Evidence from these studies suggests that the use of additional applications such macros and templates can substantially improve turnaround times, accuracy and completeness of documents generated using SR.

Discussion

The purpose of this review was to provide contemporary evidence on SR systems and their application within health care. From this review and within the limitations of the quality of
the studies included, we suggest that an SR system can be successfully implemented in a variety of health care settings with some considerations.

Several studies compared the use of transcribers to SR with human transcription having slightly higher overall word accuracy [18, 22, 33, 35] although with increased grammatical errors [29]. SR, although not as accurate (98.5 % SR, 99.7 % transcription [35]) with 10.3% to 15.2% error rates [31, 38], does deliver other benefits. Significantly improved patient outcomes such as reduced turnaround times for reporting [20, 23, 33-35] and cost-effectiveness [20, 35] have been demonstrated, however, equivocal evidence exists on improved workflow processes with Derman and colleagues finding no significant improvement [30].

Several issues related to the practical implementation of SR systems have been identified. As with any information system [39], a SR system represents the interplay of staff, system, environment, and processes. A diverse range of health professionals and support staff were included in these studies with no demonstrable differences in training or accuracy, however typists (including health professionals) who are competent and presumably fast typists have some difficulty adapting to SR systems [22] ie., more benefit is obtained for slower typists.

Also the length of transcription does seem to raise some concerns with text of 3 minutes or less recording time being problematic [22]. The nature of the information to be transcribed is also important as repetitive clinical cases frequently seen in settings such as radiology [33] or the emergency department [29], where templates or macros are easily adapted to the setting, are more likely to succeed. Applications relating to the writing of progress notes within psychiatry were limited in their success suggesting that other approaches or advances may be required where opportunities for standardised information is reduced [40].

In the majority of the included studies the reported error rates and improvements and other outcomes were achieved after only limited training was provided to participants who had no
prior experience with SR. Training delivered varied from 5 minutes [29] to 6 hours [22], but several researchers advised that either a pre-training period using any speech recognition system [22] for one month or prolonged exposure with SR (one to three months) [20] is preferred. This is confirmed by the improved turnaround times demonstrated in longitudinal studies [33].

Technical aspects of system selection, vocabulary applied, and the management of background noise and accented voices are all challenges during implementation. System selection is important with several systems available with varying levels of recognition errors (7.0%-9.1% IBM ViaVoice98 General Medicine Vocabulary to 14.1%-15.2% L&H Voice Xpress for Medicine General medicine Edition) [31] but with nonetheless relatively low error rates. Dawson and colleagues [41] noted that nurses expectations of the accuracy of speech recognition systems were low.

Accuracy also varies depending upon the vocabulary used with potential users needing to consider the appropriate vocabulary for the task— using a pathology vocabulary [18], and using a general medicine vocabulary [31]—to minimise recognition errors. For example laboratory studies varying vocabularies for nursing handover confirmed that using the nursing vocabulary was more accurate than using the general medical vocabulary in the Dragon Medical version 11.0 (72.5 % vs. 57.1 %) [42].

Most contemporary SR systems have advanced microphones that have noise cancelling capacities that allow for SR systems to be used in noisy clinical environments [18, 36]. SR systems now accommodate some accented voices such as Dragon Medical™ providing accented voice profiles, for Australian English, Indian English and South East Asian English [40]. Finally the use of standardised terminology is recommended such as the Voice Recognition Accuracy standards- by the National Institutes of Standards and Technology [22] when reporting study outcomes.
Limitations of the study

Whereas every endeavour was made to optimise inclusivity, the heterogeneity of the studies made comparative analysis and synthesis of the data challenging. The studies included in this review represent comparative designs or descriptive evaluations and only further rigorous clinical trials can confirm or refute the findings proposed here. A thorough examination of the cost benefits of SR in specific clinical settings needs to be undertaken to confirm some of the economic outcomes proposed or demonstrated here. The focus on patient turnaround times in reporting of radiographic procedures or assessment within the emergency department has the potential to increase patient flow and reduce waiting times. Additionally, SR has the potential to automatically generate standardised, terminology-coded clinical records and dynamically interact with clinical information systems to enhance clinical decision-making and improve time-to-diagnosis. Taking into account these areas in future evaluations will allow for a more comprehensive assessment of the overall impact that SR systems can have on quality of care and patient safety, as well as efficiency of clinical practice. We acknowledge the importance of publication bias relating to non-publication of studies or selective reporting of results that may affect the findings of this review.

Conclusions

SR systems have substantial benefits but these benefits need to be considered in light of the cost of the SR system, training requirements, length of transcription task, potential use of macros and templates, and the presence of accented voices. The regularity of use enhances accuracy although frustration can result in disengaging with the technology before large accuracy gains are made. Expectations prior to implementation combined with the need for prolonged engagement with the technology are issues for management during the implementation phase. The improved turnaround times of patient diagnostic procedure reports or similar tasks represent an important outcome as it impacts on timely delivery of
quality patient care. The ubiquitous nature of SR systems within other social contexts will guarantee improvements in SR systems (software and hardware). The availability of applications such as macros, templates, and medical dictionaries will increase accuracy and improve user acceptance. These advances will ultimately increase the uptake of SR systems by diverse health and support staff working within a range of healthcare settings.

**List of abbreviations**

ChT = Charting Time

ED = Emergency Department

eHealth = Health Informatics

HT = Human Transcription

PACS = Picture Achieving and Communication Systems

RCT = Randomised Control Trial

RIS = Radiology Information System

RP = Report Productivity

RTT = Report Turnaround Time

SCR = Speech Contribution Rates

SR = Speech Recognition

ST = Speech Technology

TT = Training Time

WRR = Word Recognition Rate.

No conflict of interest has been declared by the authors. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

The author(s) declare that they have no competing interest.

**Funding Statement:** Funding for this study was provided by the University of Western Sydney.
NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program. NICTA is also funded and supported by the Australian Capital Territory, the New South Wales, Queensland and Victorian Governments, the Australian National University, the University of New South Wales, the University of Melbourne, the University of Queensland, the University of Sydney, Griffith University, Queensland University of Technology, Monash University and other university partners.

Authors’ contributions

MJ: conception, design, acquisition, analysis, interpretation of data, drafting and revising of intellectual content, final approval.
SL: acquisition, analysis, and interpretation of data, drafting and revising of intellectual content, final approval.
VL: acquisition, analysis, interpretation of data, drafting and revising of intellectual content
PS: analysis, interpretation of data, drafting and revising of intellectual content, final approval.
HS: design, analysis, interpretation of data, drafting and revision of intellectual content, final approval.
JB: design, analysis, interpretation of data, drafting and revision of intellectual content, final approval.
LD: design, drafting and revision of intellectual content, final approval.

Author’s information

MJ Associate Dean (Research) Faculty of Health Sciences Australian Catholic University, Previously Professor of Nursing, School of Nursing & Midwifery, University of Western Sydney and Director, Centre for Applied Nursing Research (a joint facility of the South Western Sydney Local Health District and the University of Western Sydney), Sydney Australia. Affiliated with the Ingham Institute of Applied Medical Research.

SL Previously Research Fellow, Centre for Applied Nursing Research (a joint facility of the South Western Sydney Local Health District and the University of Western Sydney), Sydney Australia.

VL Research Assistant, School of Computing, University of Western Sydney, Sydney, NSW, Australia.

PS Research Nurse, University of Western Sydney, Sydney, NSW, Australia.

HS Senior Researcher, NICTA, Adjunct Research Fellow, The Australian National University, College of Engineering and Computer Science, Professional Associate, University of Canberra, Faculty of Health, and Adjunct Professor, University of Turku, Department of Information Technology, Canberra, ACT, Australia.
JB Senior Lecturer, University of Western Sydney, Sydney, NSW, Australia.

LD Associate Professor, University of Wollongong, Wollongong, NSW, Australia.

Funding for this study was provided by the University of Western Sydney.


Figure 1: Selection of studies for the review

- Total articles originally retrieved ($n = 1,730$)
- Papers retrieved for detailed examination ($n = 1,091$)
- Number of duplicates removed ($n = 639$)
- Full articles retrieved for further examination ($n = 18$)
- Papers not relevant after review of abstracts ($n = 1,073$)
- Papers not meeting one or more of the inclusion criteria (not including health professionals, not in the health domain, no evaluation) ($n = 5$) + Additional SR experimental study (government report) was added. ($n = 1$)

Papers included in review ($n = 14$)
Additional files provided with this submission:

Additional file 1: SR Table 1_Search strategy_revised.docx, 16K
http://www.biomedcentral.com/imedia/2752508414105961/supp1.docx

Additional file 2: SR Table 2_Quality scoring_Final.docx, 25K
http://www.biomedcentral.com/imedia/1265970037141059/supp2.docx

Additional file 3: SR table 3_Final_MJ_170814.docx, 35K
http://www.biomedcentral.com/imedia/9343955431410598/supp3.docx

Additional file 4: PRISMA Checklist - completed.doc, 63K
http://www.biomedcentral.com/imedia/5913108961438742/supp4.doc