

Listening to what is said – transcribing what is heard: the impact of speech recognition technology (SRT) on the practice of medical transcription (MT)

Gary C. David¹, Angela Cora Garcia¹, Anne Warfield Rawls¹ and Donald Chand²

¹Department of Sociology, Bentley University, Waltham MA, USA

²Department of Information and Process Management, Bentley University, Waltham MA, USA

Published in *Sociology of Health & Illness* (2009) Vol. 31 No. 6, pp. 924–938

Abstract

Medical records have become central to nearly all aspects of healthcare. However, little research exists on their creation. Using data from an ongoing ethnographic study of healthcare documentation production, this paper examines the process of medical record creation through the use of speech recognition technology (SRT) and subsequent editing by medical transcriptionists (MTs). Informed by ethnomethodology (EM) and conversation analysis (CA), the results demonstrate the professional knowledge involved in the work of medical transcription, which includes a combination of skilled worksite practices and an orientation toward the social order properties of recorded dictation. Furthermore, we examine how the advantages and limitations of SRTs can impact the work of transcription. We conclude with strategic recommendations for using SRTs to support medical records production and recommend against total automation.

Keywords: medical transcription, health information systems, medical records, speech recognition technology

Introduction

Automation and information technologies are increasingly being applied to all aspects of healthcare with the aim of lowering costs and improving the quality of treatment. Part of this effort is the projected \$10 billion that will be spent over the next five years toward creating a US-based national health information network, with the expectation that the investment in ‘electronic health records and new technology will reduce errors, bring down costs, ensure privacy, and save lives’ (speech by President Barack Obama, February 24th, 2009). At the same time, as Mark Chassin, President of The Joint Commission (a non-profit healthcare accreditation organisation), notes ‘Patient safety can be jeopardised when technology is not carefully planned, integrated, fixed, or updated’ (Roop 2009: 11). Thus, while technology has the promise of improving healthcare, that its application will automatically result in doing so is not a foregone conclusion.

With this growing interest in using automation and information technology in healthcare has come emergent research on its actual use, focusing on tools such as speech recognition (Shneiderman 2000, Borowitz 2001, Mohr et al. 2003) and electronic medical records/electronic health records (Berg 1998, Clarke et al. 2001, Greatbatch et al. 1995, Heath and Luff 2000, Hing et al. 2007). This research has documented essential work practices and social competencies that need to be supported by new technologies in order for technical implementations to succeed (e.g. Greatbatch et al. 2005, Heath and Luff 2000). To date, however, these studies have focused primarily on doctors (particularly radiologists) and nurses, and not on the work practices and skilled knowledge of the MTs who have been, and in conjunction with SRTs still are, primarily responsible for the creation of medical records.

In this paper we explore how MTs do the work of editing ‘back-end’ (BESR – edited after they are dictated) speech recognition-generated documents. In doing so, we will illustrate how the work of MTs is far more complex than just typing what is spoken in voice files. Their work requires complex professionally-informed interpretive acts that in turn require sustained attention to the social order properties and content of the doctor’s dictation, knowledge of medical terms and procedures, and an understanding of interactional processes, conventions of dictating, and of producing monologic speech acts. The MTs attention to these social order properties of the doctor’s (dictator’s) speech has direct implications for understanding the impact not only of the implementation of back-end systems, but also of front-end systems (FESR – edited as they are written) in which there is no transcriptionist involved.

While such technologies are marketed as a cost-effective and timely way to produce medical records, they may actually slow down the work and involve the potential risk of removing third-party quality assurance from the process. We conclude that, while new technologies have a role to play in the creation of healthcare documentation, such production technologies (and their implementation) should in most instances focus on facilitating and supporting, rather than replacing, the work of medical transcription done by skilled human workers.

Medical transcription, speech recognition, and record production

In medicine, the patient medical record has become a critical component of nearly all aspects of healthcare. Medical records emerged to support specialisation and co-ordination in medicine, making information about patients available to the increasing number of personnel involved in treatment and payment (see Timmermans and Berg 2003). This was not always the case – as the

country general practitioner had little use for such extensive documentation. As healthcare becomes more complex, the importance of medical records continues to increase. In facilitating co-ordination between groups, medical records operate as important ‘boundary objects’ (Star 1999, Star and Griesemer 1989) that cross organisational boundaries and can be accessed by a variety of users, including doctors, reimbursement agents, insurance companies, legal professionals, medical researchers, billing coders, audit contractors, and the patient. Given this diverse and distributed community of use, the comprehensive detail and integrity of the record has become very important.

Records are not simply reports of ‘what happened’, but accounts created for social purposes and uses, situated in various institutional contexts and without which neither co-ordinated action, or even compensation for services, may be possible. Garfinkel (1967:186-207) noted in his study of record-keeping practices in a psychiatric clinic that, while records are created by individuals, they must be recognisably valid as competent accounts in Speech recognition technology and medical transcription organisational terms. He called this ‘institutional accountability’. Similarly, medical records are not just for the originating doctor, but must also be recognisable and meaningful to other healthcare professionals. Formatting such documents according to widely recognizable standards – in institutionally accountable terms – is essential to their usability and validity. While a limited number of people may be present for a given medical encounter, the number of potential ‘witnesses’ to the event increases when the encounter becomes an accessible record.

Despite the centrality and importance of medical records, and the growing interest in them, relatively little research has been done on the process of their creation. Most attention has been given to the new technologies, while the skilled practices and social competencies of MTs have received relatively little attention. To the extent that the work practices of MTs add significant value to the medical record, systems that replace them could have a negative impact – unless they take the competencies of MTs into account. That the vast majority of MTs are women, at a time when the pay and status of women are still unequal, may have contributed to their current invisibility. However, their actual visibility has also changed. In the past MTs worked directly in the hospital or clinic setting, sometimes sitting in operating theatres taking dictation during surgeries. This is however no longer the case for most MTs. New technologies (improved information networks, computers, digital voice files) have resulted in most MTs telecommuting (with approximately 72 per cent of MTs worldwide working from home (David et al. 2008)). Thus, their figurative invisibility has now become literal in their absence from the healthcare setting.

Arguably the most significant change to the work of creating medical records has been the introduction of SRTs to create a written transcript of a doctor’s dictated notes on a medical encounter. There are two ways that SRTs perform their role: synchronous recognition (‘front-end’ speech recognition — FESR) or asynchronous recognition (‘back-end’ speech recognition — BESR). Dictation typically occurs when doctors orally record (through a telephone, cell phone, microphone, or into a handheld tape recorder) a version of what occurred in the medical encounter, including any supplemental ‘relevant’ information from past encounters, lab reports, etc., which they decide is important for patient care (and billing). With synchronous recognition (FESR), speech is immediately transformed into text, which can be viewed by the dictator as s/

he speaks. With asynchronous recognition (BESR), the digitised voice file is processed after the dictation is completed, and then transformed into a written version.

Both methods require editing. The difference is that with FESR editing is possible at the same time of the dictation, while with BESR editing is only possible after the entire record has been dictated. Another important difference lies in who does the editing. With FESR, the original dictator (often a doctor) is typically the one to edit and check the document for accuracy; in BESR it is typically the MT who performs the editing function by listening to the original dictation voice file and comparing it to what was produced by the SRT system (while also checking for dictator accuracy). Thus, in BESR, the MT performs two checking functions: (1) checking the accuracy of the SRT, and (2) checking the accuracy of the dictator.

Technology vendors have advocated both synchronous and asynchronous uses of SRTs for the production of cheaper, faster, and more accurate medical records. The extent of these claims varies by vendor, and their comparability is complicated by the different methodologies used by each company to justify their product (Speech Recognition Adoption 2008). While companies differ in whether the technology is offered as a replacement of medical transcription or as a tool for assisting the process, a consistent claim across the speech recognition industry is that SRTs can reduce costs due to faster turn-around times of medical documentation, higher efficiency, and increased accuracy. One hospital is reported (Wright et al. 2007) to speak of eliminating all transcription costs. An article entitled 'Less paper, less fuss, better patient care' reported greater efficiency and better patient care as a result of implementing health information technology (Schock 2007). SRT vendors (e.g. M*Modal, eScription, and Nuance) provide case studies that support the claim that their products will deliver on their promises, presenting SRT either as an important transcription aid, or as a complete replacement of those who do transcription.

Despite these promises, the literature reveals that results of SRT implementation have been mixed. Conn (2005: 38) reports that while 'technology is supposed to make medical transcription obsolete...the demise of the medical transcription industry is greatly exaggerated'. This is due to the lack of accuracy found in SRT systems. Because of problems with accuracy and ease of use, 'experts confirm that outside of radiology and pathology, relatively few physicians have adopted the technology' (Goedert 2006: 44). While SRT has its place, van Terheyden (2005: 45), the Chief Medical Officer of a SRT company, notes that 'it is neither a panacea nor the Holy Grail'.

Part of the problem stems from the way in which SRTs are able to perform their work. SRTs generally treat transcription as a process independent of social and professional competencies. In order to produce a clean transcript SRTs treat hesitations, repeats and many other social order properties of speech as 'disfluencies' (see Extract 1). In cases of ambiguity SRTs do not leave out or flag terms as problematic. Instead, they 'choose' a word, often based on 'mathematic probabilities of when and how often words will appear in a particular context' (AHIMA 2003: 1). SRTs cannot create order in a file if the dictator has failed to record explicit and clear directions for this. Similarly, they cannot correct dictator errors, correct punctuation – or treat a repeat as a correction –without direction.

The work of doing medical transcription involves a complex set of tasks (see David, Chand and Garcia 2008, Garcia et al. 2008). MTs not only type what is dictated. They verify the spelling of

medical terms and orient complex social order properties in the dictation, as well as recognise repeats, hesitations, and corrections in the monologue. They also flag critical medical errors and organise the transcript to meet the requirements of an organisational and legal record. In addition they can discern voice inflections for punctuation, convert spoken language into grammatical written language, filter background noise and distractions in the voice file, and check specific elements in the dictation against the larger context of the record. In cases of ambiguity they are trained either to omit or flag the questionable item. MTs are able to accomplish these tasks not only because of their training¹ in anatomy and physiology, pharmacology, legal issues, laboratory testing, and a range of medical specialisations, but because of their knowledge of mundane procedures and routine practices of interaction and spoken language. Through the application of this combination of professional and mundane social knowledge, the transcriptionist generates a record that complies with medical and legal standards – rather than just reproducing a sound file in written form. Transcription involves essential knowledge work based on social practices and is not just manual labour (i.e. typing).

The extent to which work is perceived as largely ‘manual’ or ‘knowledge’ work often is important when determinations are made whether technology can or should replace and/ or restructure work. Work that is considered more manual than knowledge based is typically at greater risk of being impacted by technology implemented as ‘labour saving’. In fact, the most seemingly menial jobs can require extensive professional sense-making practices. As Rose (2004: xxvi) notes, ‘In the rhetoric of the “new economy”, communication skills, general problem-solving skills, or the ability to work in teams are privileged, and more likely to be treated as knowledge work, while more specific mechanical skills – associated with Speech recognition technology and medical transcription conventional blue-collar work – tend to be perceived as less valuable’. Research examining workplace practices has demonstrated that much knowledge and professional sense making goes into work that is deemed ‘routine’ and ‘mundane’. For instance, Suchman (2000) in her study of legal secretaries found that informed judgment had to be exercised in examining and coding documentation related to a legal case. As Whalen, Whalen and Henderson (2002: 239) note, ‘the traditional topics of “work routines” and “routinisation” need to be respecified in order to take into account how any “routine” is a contingently produced result.’ In the end, attempts to ‘save labour’ through technological standardisation can have the impact of actually impeding work due to an interference with essential work skills, practices and routines.

The implementation of labour-saving technology, and more general standardisation of work, has also been linked to a deskilling of workers. Braverman’s classic studies (1955, 1974) of automation questioned the extent to which technology contained the promise for a better society or sowed the seeds for social problems. Leo Marx (1987) questioned whether technology brings progress, or whether people are becoming increasingly disenchanted with the unfulfilled promise, and sometimes unethical application, of technology. Many studies address the impact of technology on work and deskilling (Garson 1988, Zuboff 1988, Orlikowski 1992, Harper et al. 2000). As Castells (1996: 263) notes, ‘The effects of these technological changes on office work are not yet fully identified, because empirical studies, and their interpretation, are running behind the fast process of technological change’.

The impact of labour-saving technology on healthcare specifically has become an important topic for examination. Ritzer and Walczak (1988: 13) observe ‘Not only are advanced technologies coming to control physicians, they are also helping to reduce the physician’s authority over patients’. Greatbatch et al. (2005) found the implementation of a telephone triage system, while providing a certain structure to the performance of work, did not explicitly dictate (or standardise) how that work was actually delivered. Rather, nurses continued to try and exercise their professional beliefs and practices, which could pose significant challenges to technological implementation.

While there are some industry studies, however problematic, on the accuracy of the SRT, little research exists on the impact of SRT technologies on the actual work of creating medical records. To understand the impact of SRT on how MTs do their work, this paper examines the use of an asynchronous SRT system (back-end BESR) by MTs in an acute healthcare setting. We examine how the technology integrates into the existing workplace practices of MTs (part of a larger study on medical transcription and the production of healthcare documentation). While SRTs can be an effective tool, the data also reveal potential limitations and dangers of over-reliance on SRTs. The technology not only impacts on how the work gets done, but has potential effects on the quality and the integrity of the resulting medical record.

Data and methods

The research reported here is part of a larger ongoing ethnographic project investigating the production of healthcare documentation. Fieldwork has involved observations and participation with both MTs and their professional associations, and with professionals and taskforces in the SRT industry. The project is rooted in an ethnomethodological (EM) approach that focuses on ‘the ordinary ‘methods’ through which persons conduct their practical affairs’ (Lynch 1993: 5). Founded on the work of Harold Garfinkel (1967), the central argument is that all meaningful social objects and practices, including work, are constituted by social order properties and ‘depend for their coherence on constant attention to, and competent display of, shared member’s methods (ethno-methods) rather than on formal structures, or individual motivation’ (Rawls 2008: 701). In the case of MTs the voice file and its socially ordered properties – as oriented and recognised by MTs – are the focus of analysis. EM is concerned with what Garfinkel (1996: 6) calls the ‘What More’ can be understood about the production of any local order phenomena beyond that which can be rendered through formal analytic approaches. Our research seeks to uncover the details of ‘transcription-at-work’.

Along with attention to the detailed production of work, project data also included an online survey of MTs (n=3807), 33 interviews with persons in the MT industry, eight MT focus groups, and fieldwork and observations of MTs and dictators using speech recognition technology, doctors using front-end SRTs, and audio and video of the process. In addition, the first author has been involved as an observer and participant on numerous SRT industry committees related to quality assurance, performance metrics, and the adoption of SRT. All subjects were made aware of the project’s goals, and interviews and focus group participants signed participant confidentiality statements.

To demonstrate the challenges associated with SRT transcription and the MT work of editing that transcript, we conducted a single case analysis of an MT, working at a regional New

England hospital, who was using a back-end speech recognition engine to edit a physician dictation. The dictation was made by a neurologist recording an encounter with a patient hospitalised for head pain.² As in studies of conversation using a single episode of interaction (Schegloff 1987), we examine a single case to understand in detail the features of the human machine interaction that occurs between the dictator, the technology, and the MT. To do so, we transcribed a version of the dictated voice file (VF) that attempted to capture as many details of the sound file as possible – with a focus on preserving elements of its character as a sound file – without reducing it all to words, punctuation and other social conventions of speech. The entire 19 minute 32 second dictation was transcribed using conversation analytic³ (CA) notations (see Jefferson’s transcription system in Atkinson and Heritage 1984). As in CA generally, punctuation in our VF transcript records punctuation marks where they can be clearly heard – rather than where they belong grammatically. The VF transcription was then analysed alongside the SRT-generated draft report (before MT editing), and the final edited medical record (the creation of which by the MT was observed in the worksite).

The creation of medical records is a social act (see Heath 1982), and therefore can be analysed as such. By observing in close detail the way the SRT ‘hears’ and transforms spoken words to text, and performs the formatting instructions included in the dictation, we can examine how the SRT performs. Similarly, as we analyse the MT’s correction of the SRT’s transcript in light of the doctor’s dictation, the skills, knowledge, and practices used by MTs become visible.

Examining a single case of making a medical record

Natural features of dictation and speech recognition

The ability to recognise and treat speech disfluencies (such as hesitation markers (e.g. ‘uh’), pauses, drawing out of speech, ‘tisking’ sounds, changes in pitch, emphasis, and repeats / stutters) as meaningful social order properties of speech (i.e. indicating correction, punctuation, reordering) is a social skill necessary for communication. The sequential Speech recognition technology and medical transcription positioning of these features can impact on how turns of talk are understood and how words are heard.

While the capabilities of SRT systems vary, one common feature is filtering out speech disfluencies – eliminating them from the transcript altogether. While some speech disfluencies are indeed unnecessary, others make a significant contribution to the sense of talk. For instance, the SRT can inadvertently omit the indefinite article ‘a’ when it is pronounced as ‘uh’ rather than ‘_a’. The SRT sometimes misses the significance of error repair (when the dictator repeats a word to correct an error in dictation). SRT technology is not always successful in differentiating between consequential and inconsequential disfluencies, tending to filter too much. SRT are marketed as having the ability to ‘learn’ from past mistakes. But as Shriberg (2001: 153) points out ‘speech recognition models are often trained on read or highly constrained speech’. Subsequent corrections by the MT only successfully ‘train’ the SRT when the corrections are consistent. If the same mistake is corrected in the same way a number of times, then the SRT will ‘learn’ to produce the corrected term whenever it identifies the original term in a sound file. Both this ability and the need for the MT to attend to it introduce problems of their own

Extract 1 demonstrates some of these issues.

Extract 1: Lines 20–21 of Dictation (all line #s refer to VF master transcription)
(VF is our transcription, followed by the SRT and MT versions)

VF (0.2) no history of uh seeing zig zag lines, white or dark spots in front of her eyes in front of his eyes, (0.4) .h h no uh photophobia (0.3)

SRTno history of seeing Z.
5. Dark spots in front of her eyes in from his eyes. No photophobia.

MT ...no history of seeing zig/zag lines, bright or dark spots in front of his eyes.
There was no photophobia...

There are a number of points to note here. The SRT drops the ‘uh’ before ‘seeing’, as it is supposed to. But, then it also removes part of the description of symptoms, rendering ‘zig zag lines, white or’ as ‘Z. ((new line))’. The SRT then treats some part of the disfluent speech as punctuation and begins a new sentence and a new line – with an additional space – with the terms ‘5. Dark’. This creates format problems in the document. The SRT also misses a common dictator error; the doctor confusing and then subsequently correcting the gender of the patient. The doctor first says ‘her eyes’ then corrects this mistake to ‘his eyes’. The SRT changes the repeated words ‘in front’ before ‘her eyes’ and ‘his eyes’ to ‘in from’ when they occur the second time. Not picking up on the error repair format, the SRT adds confusion and error to the transcript. Lastly, in recognising ‘No photophobia’ the SRT begins a new sentence as indicated by the dictator, but does not correct it to a complete sentence. In so doing it performed appropriately, but the resulting record does not look as competent as it might.

The MT, by contrast, is able to recognise the error correction format, and clean the draft: treating ‘zig zag’ as a meaningful term, not as a repeat that should be eliminated; retaining the words ‘white or’ instead of eliminating them; and turning the words ‘no photophobia’ into a grammatical statement. The MT, orienting the social order properties of language and making use of both professional training and ordinary human speech competencies, is able to produce a more accurate, grammatical and organizationally presentable document.

Extract 2 displays similar problems with disfluency resulting in format issues.

Extract 2: Lines 33–35 of Dictation

VF .h next line uh (2.0) h uh next line um (1.5) past medical history (5.0) .hh unremarkable (0.2) except number one. number two, .hh he describes that accidentally he shot his uh (0.2) .h left hand

SRT Knee and past medical history.

Unremarkable except

- 1.
2. He describes that he does extend he short his left hand...

MT **Past Medical History**

1. Unremarkable.
2. He describes that accidentally he shot his left hand...

First, the doctor provides the instruction ‘next line’, which the SRT follows. But, when the doctor repeats the instruction, the SRT renders the second instance of ‘next line’ as ‘Knee’, thus not only missing the repeat – but rendering the repeat as a section header – but without proper section header format (bold) and content. These headers are important as they often guide the reading of the record when readers are searching for information.

The doctor then dictates the patient’s medical history out of order. After the direction for a major header ‘past medical history’ he begins the medical history without a number sequence and then makes the correction. The utterance ‘.hh unremarkable (0.2) except’ is followed by the words ‘number one. number two’. The SRT does its job of recognising what was said, but does not correct it. The SRT hears a command for ‘number one’ followed immediately by a command for ‘number two’ – and types the transcript as such. Finally, while the SRT is able to filter out various hesitation markers and in-breaths that are not relevant to the record, it misses the essential information about the gun accident, rendering it as ‘he does extend he short his left hand’.

The MT formats the header to conform with expectations regarding document form, and then orients the number sequence to the requirements of a medical history. The MT also removes the word ‘except’ from the record and corrects the description of the gun accident. Generally, the MT applies professional knowledge of the information that needs to be contained in the record, knowledge of the preferred format, and mundane reasoning skills regarding social order properties of the dictation, to make the final version of the record.

Doctors rely on MTs to ‘fill in the blanks’ and understand how the record should look. SRTs do not possess the same ability. While they can often recognise words, they are not able to orient to social ordered properties of those words (the properties of a list, the properties of a correction), nor can they orient to expectations regarding document format without direction. MTs, employing social competencies, are able to treat the dictation as an ‘intendedly unified object’; a social gestalt, and thus to compose the information into a complete, unified and organisationally accountable social object.

Punctuation, instructions, and dictation

While some dictators provide directions for punctuation, doctors typically rely on MTs to provide punctuation as needed to make the record a readable and socially and organisationally competent document. To accomplish this, MTs rely on their hearing of the dictator’s intonation, their understanding of the content and the social order properties of speech, and their knowledge of how the finished document should look. SRTs sometimes add punctuation at appropriate points, but problematic placements and elisions of punctuation often occur. SRTs work from a rule-based application of grammar, whereas MTs have a social and situated understanding of sense-making practices.

In Extract 3 the SRT's use of punctuation creates errors in the transcript.

Extract 3: Lines 23–24 of Dictation

VF (0.7) at thuh ee ar he had head cee tee scan which was reported as unremarkable .hh followed by a spinal tap which revealed uh .h u:h unremarkable u:h cee ef es (0.4) s: studies. (0.8) .h next line since his admission...

SRT Sensory ER he had a head CT scan which was reported as unremarkable followed by a spinal tap which revealed unremarkable CSF. Studies since his admission...

MT In the Emergency Room he had a head CT scan which was reported as unremarkable followed by a spinal tap which revealed unremarkable CSF studies.

Since his admission....

The SRT adds the word 'sensory' at the beginning of the first line. Then at the end of the second line puts a full stop after CSF and begins a new sentence with 'Studies.' This breaks up a test result 'CSF studies'. In the dictation, 'studies' has audible 'period intonation', making it hearable as the last word in the utterance. The SRT's failure to insert a 'period' (full stop) after 'Studies' is odd given that it 'hears' and implements the 'next line' command which follows this word. A line break in the middle of a sentence is unlikely. The MT hears 'CSF' as belonging with 'studies', and then hears 'next line' not only as an instruction to begin a new line, but also to insert a space creating a new paragraph. In addition to correcting the beginning of the utterance and changing it into a competent sentence, the MT performs two analytical functions that the SRT does not: hearing the punctuation intonation in the doctor's voice, and applying formatting commands with an orientation toward overall document formatting requirements.

The issue of misplaced punctuation may seem minor. However, any error in a record can raise suspicion regarding the accuracy of its content. As errors increase, the readability of the record can go down. In a narrative example reported by MTs at the same facility the authors were told that a physician dictated 'Since she was transfused she notes she definitely feels stronger'. However, the SRT system rendered, 'Since she was transfused she notes that she does not feel stronger' (emphasis added). Such an error is particularly worrisome because there is nothing noticeably wrong with the sentence itself to indicate there might be a problem, yet it inaccurately portrays patient condition.

Furthermore, patient care is just one use of medical records, and errors that may not have consequences for patient care can have significant repercussions in a legal setting, or in billing. Consequently, the organisational goal is to have a record with few, if any, errors.

Not marking erroneous recognition: numbers and lab values

When an MT cannot decipher an utterance, she is trained to leave it blank or otherwise 'flag' it as unclear. There are two ways the SRT approaches sounds it treats as speech (not disfluencies) but that are difficult to decipher. The first is not transcribing them and marking them

nontranscribable with the word ‘<skip>’. The second is to transcribe the speech incorrectly. The SRT does not currently seem to have the capability to mark a term it does transcribe as tentative.

SRTs sometimes have problems with numbers, whereas identifying relations between numbers in combination constitutes an ordinary human competency for the MT.

Extracts 4 and 5 demonstrate this:

Extract 4: Lines 50–51 of Dictation

VF ...pulse is fifty eight: .h temperature ninety seven point four

SRT ...pulse is 50, 8, temperature 97.4

MT ...pulse is 58. His temperature is 97.4.

Extract 5: Lines 73–74 of Dictation

VF () is one point zero two (and calcium) is ten point two and uh (5.0)

SRT ...creatinine is 1.0 and calcium is 9.2...

MT ...creatinine is 1.02 and calcium is 10.2.

In Extract 4, the doctor’s emphasis on the word eight (indicated by underlining in the VF transcript) may have caused the SRT to hear it as separate from the number begun as ‘fifty’. However, in Extract 5, the SRT transcribes a creatinine level of 1.02 as ‘1.0’ and a calcium level of 10.2 as ‘9.2’. Elsewhere in the record, the SRT transcribed a hematocrit level of 41.3 as ‘41.37’ and lymphocyte 7% as ‘lymphocytes seven was sent’. These are potentially significant errors for test results that the MT was able to notice and correct.

Incorrect lab values can cause major complications for patient care. It is common for dictators to speak rapidly through lab values, and MTs who have access to labs or a patient’s medical history can use that information to clarify unintelligible dictation. A SRT does not have that ability, and therefore relies solely on what the dictator says.

Furthermore, MTs are trained to know the normal range of lab values. If lab values do not match the rest of the record (as when abnormal values are given for an otherwise normal patient), MTs can double check this information (when they have access), or at least flag the potential source of error.

Correcting the dictation

A number of factors work against the goal of creating an intelligible and presentable medical record. The chief challenge can be the dictator. A 2004 study by AHDI found that 27 per cent of errors in medical records were linked to dictator mistakes: including reversing left and right, changing genders, dictating the wrong lab values, and confusing patients and patient numbers. The MTs interviewed describe dictation marred by doctors eating and drinking while dictating, talking on cell phones, driving cars, exercising on treadmills, or transcribing in noise-rich environments.

Correcting the doctor's dictation is one of the most important functions of the MT – and one in which both professional training and social competence play a large but often unacknowledged role. Working with SRT systems can interfere with this function if the MT becomes focused on correcting the SRT's mistakes, and also 'training the system,' and not on the dictator's mistakes. Training the SRT basically means that the system can become 'smarter' when a word is corrected and the system 'learns' to recognise the dictated word in relation to the correction. However, this only works if the correction is consistently applied in the same manner. For an MT, accuracy can become a matching process between the SRT transcription and the dictation. The question of whether the original dictation was in fact correct can get lost. MTs have told a number of stories about how the SRT typed the name of a drug correctly, but the doctor had dictated the wrong drug. The MT noticed because the condition being described would not be treated by the drug named in the dictation. When the MT is distracted from the task of checking the doctor's dictation for mistakes in order to verify and check what the SRT has transcribed and correct it consistently, medically important dictator errors can be missed.

In addition to these difficulties, our survey and ethnographic data reveal that problems can result because the technology changes how the work is done. In interviews, MTs have talked about the brain working differently when transcribing dictation from when editing SRT generated documents. One MT explained it as 'funneling the voice file through your fingers (when transcribing from a voice file) versus through your eyes (when editing a SRT document)'. Others speak of 'following what the technology produces versus following their own knowledge'. As a result, when working from an SRT-generated transcript, MTs can end up paying more attention to whether the draft matches the dictation than focusing on whether what the dictator said makes sense. This divided attention can result in missing important errors in the dictation.

The response to the new technologies by MTs is mixed. Some MT instructors worry that new MTs will be intimidated by the technology and consequently less likely to correct what the 'machine tells them'. Some MTs report that they like using SRT and that it increases their productivity. However, editing does not always 'save labour' and MTs also report increased exhaustion because of the intense focus necessary to read text while listening to voice files. One report indicated that the focus required was similar to driving in a blizzard, in which constant attention is needed. Our survey revealed mixed enthusiasm for SRTs, with about equal percentages of those who responded to a question regarding experience with SRTs indicating that they could 'take it or leave it' (30%) or saw it as 'helpful in making me more productive' (27.9%).

It is clear that SRTs have a place in creating healthcare documentation; the task now is to determine what that place is.

Conclusion: automating versus supporting medical transcription

Whalen and Vinkhuyzen (2000:93), in their study of the use of expert systems in technological diagnostic work, argue that the 'expert system design should be founded upon the demonstrably indispensable aptitude and competence of users, and more responsive to the socially organised features of their activities'. While expert systems are often claimed to make experts out of non-experts through the extraction of knowledge from actual experts, such systems do not in fact

accomplish this goal (see Whalen 1995). It is not that expert systems have no use, or do not work. But, their relevance is situated and involves essential social competencies. It is important to identify the conditions under which they do work.

Because of their limitations, technological implementations aimed at the production of medical records will need to take into account the abilities, work practices and particularly the social competencies regarding language use and interaction of those who are involved in their creation. In writing about the problematic nature of SRT-generated transcripts, van Terheyden (2005: 42) comments that '[w]ritten language may need punctuation – commas, periods and quotation marks – according to strict rules that are not obvious in speech and are difficult to infer'. The MT is attuned to this. The medical record is also expected to exhibit an institutionally accountable/expected form. As doctors do not always dictate in an orderly manner, it often is up to the MT to group, separate and order entries to create form and clarity in the record. Doctors can be repetitive, add the wrong punctuation, speak in incomplete sentences, leave out information, and misspeak. In every event, the MT must decide what 'makes sense' and how it should appear. The SRT lacks this ability. The ASTM International Guide (2004: 4) to SRT products notes, 'SRT will not overcome dictation errors, improper grammar, incomplete or disorganised dictation, or incorrect punctuation'. As we have shown the problems that result are not small.

Healthcare professionals are beginning to recognise the limitations of systems such as SRTs, and the need for situated design and strategic applications of the technology that take the social properties of work and social expectations with regard to medical records into account. As Coiera (2000: 279) notes, 'To create processes and technologies that support the communication space, we first need to characterise the activities that occur within it and understand where improvement is needed'. Similarly, Shneiderman (2000: 65) observes that as SRT success stories start to increase, 'designers should conduct empirical studies to understand the reasons for their success, as well as their limitations, and their alternatives'. Or, as Greenbaum and Kyng (1991: 2) state, 'Computer systems are tools, and need to be designed to be under the control of the people using them'. While 'speech recognition is not, in and of itself, the final solution in clinical documentation' (AHIMA 2003: 1), it will have a place and function in creating medical records.

Our research has shown specific ways that the skilled work practices of MTs add value to the production of medical records. Furthermore, while MTs have typically been viewed as a cost and not revenue generating, the fact remains that in the US, with its profit-driven healthcare system, the work of healthcare professionals requires adequate documentation to generate revenue. In other words, it is not the work of the doctor (or other care giver) himself or herself that ultimately results in creating revenue; rather, the documents are the things through which procedures are represented and revenue generated. Medical coders base their billing codes on these documents, and auditors use them to determine if coding is supported by the documentation. In fact, in instances where documentation does not support the claims of medical practice, payment can be withheld or rescinded. While technology has a place in enhancing healthcare documentation production, it cannot be adopted as a replacement without significant risk not only to patient care, but also to the revenue cycle of healthcare organisations.

Based on this research, a number of recommendations can be made to those exploring the implementation of SRTs. First, the social competence and professional knowledge involved in

the work of MTs is a critical link in the effort to reduce medical errors and mistakes in medical records. Any technology that integrates into that role needs to be designed to support that work. Procedural, staffing, and technological changes intended to improve the accuracy of medical records and reduce costs associated with that work should take the contribution of MTs into account as they design their systems. Secondly, SRTs can be a tool to facilitate as much as automate. As such, they may not ‘save labour’ as much as ‘change labour’ in terms of how work is done. Thirdly, understanding the medical record as a socially-constructed boundary object that is essential for collaboration across institutional domains creates a different sense of how errors should be identified and evaluated. Medical records and other socially construction information have essential social properties (Brown and Duguid 2002, Garfinkel 2008), and as such need to be considered within the context of their use.

As they currently exist, SRTs are not able to replace the skilled work done by MTs in making medical records. But, SRTs can effectively support the work done by MTs. In this way, SRT technology might be best suited to ‘facilitate’ rather than ‘automate’. Of course, such general statements probably do not encapsulate all applications of SRTs in terms of their abilities and limitations. However, in recognising the ways in which medical records are and will be used, their creation becomes less about simply automating their production, and more about creating records that: (1) are accurate depictions of medical encounters, (2) provide relevant information for situated use, and (3) display competence of care. As tools, SRTs can be advantageous, but cannot construct records with such varied uses in mind. The issue then becomes what is the best way to design and implement SRT systems so that they can facilitate the work of MTs or other back-end editors.

Address for correspondence:

Gary C. David, Department of Sociology, Bentley University, 175 Forest Street, Waltham, Massachusetts 02452, USA
e-mail: g david@bentley.edu

Acknowledgement

The authors would like to thank the anonymous persons who gave their time to this project, as well as the editors and reviewers for their comments and recommendations. We also thank members of the medical transcription and speech recognition communities for their willingness to share their insights into their industries.

Notes

1 Training programs can range depending upon the provider. However, AHDI provides a guideline for a model curriculum, which includes approximately 400 transcription hours of practice transcription, and 100 hours of actual transcription (AHDI 2008). Generally speaking, program completion can take approximately one year (depending on the program).

2 All HIPAA regulations regarding patient privacy were observed regarding removal of patient identification from the documents and audio used in this research. The Health Insurance Portability and Accountability Act (HIPAA) was enacted by the US Congress in 1996, and (among many other elements) stipulates regulations regarding the privacy of patient medical information.

3 The goal of conversation analysis is to discover the commonsense understandings and procedures people use to shape their conduct in particular interactional settings (Sacks 1984, Schegloff and Sacks 1973, West and Zimmerman 1982, Garfinkel 2006).

References

- AHIMA (2003) Speech Recognition in the Electronic Health Record Practice Brief. <http://library.ahima.org>.
- AHDI (2004) A Survey of Error Trends. <http://www.ahdionline.org/scriptcontent/Downloads/AAMTQARepor.pdf>
- AHDI (2008) The Model Curriculum for Medical Transcription (4th Edition). Modesto, CA. <http://www.ahdionline.org/scriptcontent/Downloads/ModelCurriculum.pdf>.
- ASTM International (2004) Standard Guide to Speech Recognition Technology Products in Health Care. <http://www.astm.org/Standards/E2364.htm>.
- Atkinson, J. and Heritage, J. (1984) Structures of Social Action. Cambridge, UK: Cambridge University Press.
- Berg, M. (1998) Medical work and the computer-based patient record: a sociological perspective, *Methods of Information in Medicine*, 37, 294–301.
- Borowitz, S.M. (2001) Computer-based speech recognition as an alternative to medical transcription, *Journal of the American Medical Informatics Association*, 8, 101–2.
- Braverman, H. (1955) Automation: promise and menace, *American Sociologist*, October. http://www.marxists.org/history/etol/newspape/amersocialist/amersoc_5510.htm
- Braverman, H. (1974) *Labor and Monopoly Capital: the Degradation of Work in the Twentieth Century*. New York: Monthly Review Press.
- Brown, J.S. and Duguid, P. (2002) *The Social Life of Information*. Cambridge, MA: Harvard Business School Press.
- Castells, M. (1996) *The Rise of the Network Society*. Oxford, UK: Blackwell Publishers.
- Clarke, K.M., Hartswood, M.J., Procter, R.N. and Rouncefield, M. (2001) The electronic medical record and everyday medical work, *Health Informatics Journal*, 7, 168–70.
- Coiera, E. (2000) When conversation is better than computation, *Journal of the American Medical Informatics Association*, 7, 3.
- Conn, J. (2005) Not dead yet. *Modern Healthcare*, 35, 38.

- David, G., Chand, D. and Garcia, A. (2008) 2007 survey of medical transcriptionists: Preliminary findings. <http://www.ahdionline.org/scriptcontent/Downloads/MTSurveyReport-Preliminary.pdf>
- Garcia, A., David, G. and Chand, D. (2008) The work of medical transcriptionists: sense-making and interactional competence in the production of medical records. Unpublished manuscript.
- Garfinkel, H. (1967) *Studies in Ethnomethodology*. Oxford, UK: Polity Press.
- Garfinkel, H. (1996) Ethnomethodology's program, *Social Psychology Quarterly*, 59, 1, 5–21.
- Garfinkel, H. (2006) *Seeing Sociologically: the Routine Grounds of Social Action*. Boulder, CO: Paradigm Publishers.
- Garfinkel, H. (2008) *Toward a Sociological Theory of Information*. Boulder, CO: Paradigm Publishers.
- Garson, B. (1988) *The Electronic Sweatshop: How Computers are Transforming the Office of the Future into the Factory of the Past*. New York: Simon and Schuster.
- Goedert, J. (2006) Is now the time for speech recognition? *Health Data Management*, 14, 44–50.
- Greatbatch, D., Heath, C., Campion, P. and Luff, P. (1995) How do desk-top computers affect the doctor-patient interaction? *Family Practice*, 12, 32–6.
- Greatbatch, D., Hanlon, G., Goode, J., O'Caithain, A., Strangleman, T. and Luff, D. (2005) Telephone triage, expert systems and clinical expertise, *Sociology of Health and Illness*, 27, 6, 802–30.
- Greenbaum, J. and Kyng, M. (1991) Introduction: situated design. In Greenbaum, J. and Kyng, M. (eds) *Design at Work: Cooperative Design of Computer System*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Harper, R.H., Randall, D. and Rouncefield, M. (2000) *Retail Finance and Organisational Change: an Ethnographic Perspective*. London: Routledge.
- Heath, C. (1982) Preserving the consultation: medical record cards and professional conduct, *Sociology of Health and Illness*, 4, 1, 56–74.
- Heath, C. and Luff, P. (2000) *Technology in Action*. Cambridge, UK: Cambridge University Press.
- Hing, E.S., Burt, C.W. and Woodwell, D.A. (2007) Electronic medical record use by office-based physicians and their practices: United States 2006, *Advanced Data from Vital and Health Statistics*, no.393. Hyattsville, MD: National Center for Health Statistics.

- Lynch, M. (1997) *Scientific Practice and Ordinary Action*. Cambridge, UK: Cambridge University Press.
- Marx, L. (1987) Does technology mean progress? *Technology Review*, 90, 33–41.
- Mohr, D.N., Turner, D.W., Pond, G.R., Kamath, J.S., De Vos, C.B. and Carpenter, P.C. (2003) Speech recognition as a transcription aid: a randomised comparison with standard transcription, *Journal of the American Medical Informatics Association*, 10, 85–93.
- Orlikowski, W. (1992) The duality of technology: rethinking the concept of technology in organizations. *Organization Science*, 3, 3, 398–427.
- Rawls, A.W. (2008) Harold Garfinkel, ethnomethodology, and workplace studies, *Organisation Studies*, 29, 5, 701–32.
- Ritzer, G. and Walczak, D. (1988) Rationalization and the deprofessionalization of physicians, *Social Forces*, 67, 1, 1–22.
- Roop, E.S. (2009) Healthcare technology – it’s not foolproof, *For the Record*, 21, 5, 10–11.
- Rose, M. (2004) *The Mind at Work: Valuing the Intelligence of the American Worker*. New York: Viking Press.
- Sacks, H. (1984) Notes on methodology. In Atkinson, P and Heritage, J. (eds) *Structures in Social Action: Studies in Conversation Analysis*. Cambridge, UK: Cambridge University Press.
- Schegloff, E.A. (1987) Analyzing single episodes of interaction: an exercise in conversation analysis, *Social Psychology Quarterly*, 50, 2, 101–14.
- Schegloff, E.A. and Sacks, H. (1973) Opening up-closing, *Semiotica*, 7, 4, 289–327.
- Schock, P. (2007) Less paper, less fuss, better patient care, *Health Management Technology*, 28, 62–3.
- Shneiderman, B. (2000) The limits of speech recognition, *Communications of the ACM*, 43, 63–5.
- Shriberg, E. (2001) To ‘err’ is human: ecology and acoustics of speech disfluencies, *Journal of the International Phonetic Association*, 31, 1, 153–69.
- Speech Recognition Adoption White Paper (forthcoming) Medical Transcription Industry Association and the Association for Healthcare Documentation Integrity. Unpublished Manuscript.

- Star, S.L. (1999) The ethnography of infrastructure, *American Behavioral Scientist*, 43, 3, 377–91.
- Star, S.L. and Griesemer, J.R. (1989) Institutional ecology, ‘translations’ and boundary objects: amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907-39, *Social Studies of Science*, 19, 4, 387–420.
- Suchman, L. (2000) Making a case: knowledge and routine work in document production. In Luff, P., Hindmarsh, J. and Heath, C. (eds) *Workplace Studies: Recovering Work Practice and Informing System Design*. Cambridge: Cambridge University Press.
- Timmermans, S. and Berg, M. (2003) *The Gold Standard: the Challenge of Evidence-based Medicine and Standardization in Health Care*. Philadelphia, PA: Temple University Press.
- van Terheyden, N. (2005) Is speech recognition the Holy Grail? *Health Management Technology*, February, 42–5.
- West, C. and Zimmerman, D.H. (1982) Conversation analysis. In Scherer, K.R. and Ekman, P. (eds) *Handbook of Methods in Nonverbal Behavior Research*. Cambridge, UK: Cambridge University Press.
- Whalen, J. (1995) Expert systems versus systems for experts. In Thomas, P. (ed.) *Social and Interactional Dimensions of Human-Computer Interfaces*. Cambridge: Cambridge University Press.
- Whalen, J. and Vinkhuyzen, E. (2000) Expert systems in (inter)action: diagnosing document machine problems over the telephone. In Luff, P., Hindmarsh, J. and Heath, C. (eds) *Workplace Studies: Recovering Work Practice and Information System Design*. Cambridge: Cambridge University Press
- Whalen, J., Whalen, M. and Henderson, K. (2002) Improvisational choreography in teleservice work, *British Journal of Sociology*, 53, 239–59.
- Wright, G., Wiechert, S., Marshall, S., Merkle, S. and Dinsdale, V. (2007) Integrating ED with enterprise, *Health Management Technology*, 28, 38–9.
- Zuboff, S. (1988) *In the Age of the Smart Machine: The Future of Work and Power*. New York: Basic Books.