# The maturing of automatic speech recognition in healthcare practices

Silja Vase [1]

[1] *University of Copenhagen, Karen Blixens Plads 8, 2300 København S, Denmark*

### Abstract

This thesis investigates practices in healthcare and how healthcare professionals adapt in relation to Automatic Speech Recognition (ASR) as the algorithm is employed at Danish hospitals. Previous studies expect more mature generations of the ASR algorithm to be better received among physicians [1]. Taking this expectation into account, this study analyses orthopedic physicians at two hospitals that underwent an ASR implementation. The two hospitals represent various generations of the algorithm; one of the hospitals had the algorithm implemented several years back and still experiences frequent system modifications. The other is currently undergoing the implementation of a newer version. Based on multi-sited fieldwork, an organizational study of patterns of socio-material practices, and their effects, this thesis identifies practices to assess the degree to which these adapt in the meeting with and the use of ASR.

### Keywords

Electronic Health Records, Automatic Speech Recognition, Human-Computer Interaction

## 1. Introduction

The fundamental objective of implementing healthcare technologies, such as Automatic Speech Recognition (ASR), is to enable improvement of efficiency through documentation time and accuracy and further to allow patients to access their Electronic Health Records (EHR) the same day as their medical consultations [2][3]. ASR offers an alternative input medium that claims to be faster for physicians than the use of a keyboard; however, compared to the former use of dictation, the digital transformation of practices concerning ASR is often associated with increased documentation time and outsourcing of healthcare professionals [4][1]. This thesis evolves from previous fieldwork that found ASR as a mandatory digital medium and sees an opportunity to study healthcare practices as they adapt to the phenomenon of ASR. To sum up the considerations above, this study will examine the following research questions: *How do workflows vary in relationship to Automatic Speech Recognition as it is adapted in practice and intervenes within those practices? And what values and assumptions are implicated when healthcare professionals use the technology?*

While ASR plays a central role for healthcare professionals, research remains scarce [4] and has a broad statistical angle. This thesis analyses challenges which should not be addressed solitary throughput statistical substances, why the following pages endorse a social perspective.

## 2. Automatic Speech Recognition

In short, physicians use ASR to conduct EHRs, which documents ward rounds, operations, and clinical results in general. To use the technology, they speak into a microphone connected to a computer where their speech is then translated into written text, which appears on their screen within seconds after the physicians have spoken into the microphone. They can subsequently edit the text using a

keyboard to train the algorithm for a better translation in the future. ASR is considered a mandatory medium for physicians to use when they conduct EHRs. ASR has matured over the last few years [3] while it has evolved to 'smaller' languages such as Danish [1] is spoken by nearly 6 million people. Since introducing the technology to Danish healthcare, hospitals implement ASR to enable faster completion of EHRs and to upload these in real-time [4]. ASR is thus part of physicians' daily workflow where the technology offers a replacement to the 'standard' way of entering records where they either type or dictate the records, which are subsequently transcribed by medical secretaries. Medical secretaries' role has now changed, and they have taken over other responsibilities than transcription, such as setting up physicians' ASR accounts.

Both hospitals analyzed in this thesis use Danish suppliers of the ASR algorithm developed by American based Nuance. The recent implementation is supported by a deep algorithm (Dragon) and the other by a linear algorithm (SpeechMagic). Scholars contribute to the discussion regarding the use of solutions for vocal interaction supplied by Nuance [5][6], which are supported by a deep algorithm trained by the everyday users such as Siri and Alexa. However, the Nuance supported vocal interaction ASR analyzed in this research can only be trained by the healthcare professionals working within the field of which it is used (such as orthopedic surgery) and shortens the training fundamentally compared to the others.

## 3. Related Work

ASR has been in Human-Computer Interaction (HCI) researchers' searchlight for quite some time and has matured and become more widespread in healthcare and other places. Alapetite et al. [1] covered physicians' attitudes towards ASR, which depended on how physicians considered American healthcare procedures related to EHR. The normative effect on how physicians in the study perceived the technology illustrates how a survey of ASR should not only consider the technical aspect. Other investigations concerning ASR have been surveying vocabularies as they enable the recognition rate of words spoken. However, many other factors play a role when it comes to recognition rates, such as clarity of pronunciation [7], external noise [8][9], and 'stød' [10]. Studies have thus emphasized the more technical aspects, yet, external artifacts and normative considerations interfere with how physicians perceive the technology. Since this study examines two levels of technological maturity, the algorithms' sensitivity to external artifacts may differ and, further, contribute to a different perception by physicians.

Various studies have critically inspected algorithms and how they are configured in practice by drawing on ethnographic research [11] [12][13], questioning how algorithms can be located. The studies point out an obstacle when algorithms are purely researched as mere code, since "the limits of the term algorithm are determined by social engagements rather than by technological or material constraints" [13]. Social engagements with healthcare technology often outline a meeting between users and complex technology and can lead to challenges that force users to develop alternative workflows or workarounds. Kobayashi et al. define workarounds as "informal temporary practices for handling exceptions to normal workflow" [14]. A 'normal' workflow is, however, normative to the user as the user's actions may not carry out as expected, unlike algorithms that follow the rules to conduct action or not. The use of ASR should thus be found by focusing on the actual actions made in practice. Further, little research has assessed the unveiled potential risks and benefits associated with workarounds emerging from technology implementation and the further use of these [15].

## 3.1. Practices in Theory

Considering the advised social aspects illuminated above, this study will focus on the healthcare professionals and their daily work by using practice theory. More recent accounts of practice theory have led to a renewed engagement by referring to a "practice turn" within the social sciences [16], staging that there is no unified account of practice theory [17]. The actual practices, actions, and expressions, and materiality's role should be central when analyzing the social [16] aligning practice theory with socio-technical approaches. Schatzki defines practices as "a temporally evolving, open-

ended set of doings and sayings linked by practical understandings, rules, teleo-affective structure, and general understandings" [18].

In his study of organizational phenomena, Nicolini [19] develops a site ontology that draws from the practice turn [18], presuming that all social phenomena are rooted in practices. Employing an organizational practice-based approach, this study perceives practices as defined by Schatzki and are further seen as "open-ended" [19] in an organizational environment. The perspective sees practices as continuously evolving through the use of artifacts such as technology and can comprehend not only the social practices themselves but also the technological mechanisms that engage with or support a practice. Additionally, I expect the approach to lead to an analysis that considers practices in the mentioned orthopedic departments and leads through several other sites since ASR could touch upon numerous workflows before it ends at the supplier.

## 4. Methods

Data for this study is collected during ethnographic fieldwork, which is divided into two parts. Each part extends over three months of shadowing physicians, qualitative interview [20], focus group interview [21], and observation of healthcare professionals at orthopedic departments in the Region of Southern Denmark and the Region of Zealand. To locate practices in real-time, three physicians at both hospitals will be shadowed. As the name suggests, it is a research technique that involves the action to follow a member of an organization closely for a period of time. Shadowing as a research technique is well fitted for organizational research, as it allows for follow-up questions and short interviews during an in-depth observation [22]. In addition to the fieldwork, the suppliers for both ASR algorithms are interviewed several times to grasp how the technology reflects the practices and further understand the technical factors throughout ASR's maturation.

To do fieldwork in healthcare is doing fieldwork in challenging environments, yet, to research organizational practices and how they evolve with technologies should be studied in naturalistic settings of the practices. I will here be inspired by Nicolini's [19] concept of 'zooming in on' and 'zooming out of' practice by setting myself in different observation positions during the fieldwork, where I will focus on both aspects in the foreground and others in the background. Real-time practices as a starting point will then be revisited once the data is obtained. The concept allows the researcher to track the ways in which practices are connected. The goal is to understand how practices are related in terms of time, understandings, rules, teleo-affective structure, and general understandings.

Further, the study's validity is grounded in data conducted in a specific context; it is thus central to understand in which context the field is recognized to connote the methodology or, in other words, to know "what you are looking at" in the field [23]. Since I use the concept' zooming in on' and 'zooming out of' practices, I attempt to avoid following a predetermined plan for what I assume is crucial to a certain context, but instead, understand the context as understood by practitioners. I ought to displace myself as a researcher to achieve larger or perhaps more complex contexts. To support such matters, I use fieldnotes noted in real-time and subsequently transcribe them. Once there is a break from the fieldwork, I make a short description of what has been done to interpret and perhaps find secondary elaborations. Context is thus found through an individual practice that is historically situated and determined [23]. I will find and frame the somewhat setting continuously during the fieldwork to meet the field with an overall setting and allow for it to change as the fieldwork develops.

The first part of the fieldwork was completed this summer. Three doctors were shadowed during their work in general offices, surgical wards, and the ambulatory. Healthcare professionals at the orthopedic department at a hospital in the region of Southern Denmark were observed and several interviewed. ASR was implemented two decades ago, yet, the technology still undergoes changes and updates rather frequently. In the following, I will exemplify some of the preliminary findings to illustrate the progress made in this thesis.

## 5. Progress Made

In order to use ASR when conducting EHRs, physicians need to have an account connected to the regional EHR system. Secretaries set up the accounts in collaboration with the physicians who are then able to use the technology. The setup is rushed due to little time set aside for each physician, and if any complications prevent the setup from being done in time, the account cannot undergo complete training. During an observation in the field, some complications arose, and accounts were thus partially accumulated with a lower likelihood of recognizing words than other accounts. The complications arose when a physician who spoke with a strong accent should set up her account. The algorithm builds on Danish as a natural language; however, not all physicians use the same accent or even the same syntax, resulting in a distorted recognition of words from one account to another since they do not represent the standard accent syntax. Due to conditions such as language barriers or uncompleted training, some physicians need to train their accounts more than others to reach a recognition rate corresponding to other physicians.

Individual recognition rates lead physicians to individual results, use, and description of ASR. During the fieldwork, several ways of conducting EHRs using ASR were observed. An example is during consultations with patients in the ambulatory, where Jørgen, a physician, uses ASR after every consultation while the patient is still in the room and can follow the conduction. Jørgen speaks whole sentences and correct unrecognized words using the keyboard after they appear on the screen. If the system works slowly, and he has to wait more than a few seconds after a sentence is spoken to see the translated words, he takes a break by getting a cup of coffee or looks through other patients' records. He describes the technology to change according to time of the day:

"I experience a delay during the afternoons and especially during nighttime. Perhaps my voice changes but I think the system is rebooted or something. It is always slower during the nighttime" (Jørgen, Physician)

Another physician, Peter, uses ASR after the patient has left the room and speaks in interrupted sentences. He articulates less patience by refreshing the system often or restarts the whole computer if the technology recognizes words slower than after just two seconds or 'freezes'. Suppose the system does not improve after refreshing the EHR program or restarting the computer. In that case, Peter starts using the keyboard as input but switches over to handwritten notes while complaining about the system:

"It never works. Why should I use this as a physician? It never works. Now I have to wait until it works again, and I have to record the whole days' work" (Peter, Physician)

When speaking about ASR, the physicians consider the technology differently; one considers it to change due to technical demands that are out of his hands. The latter takes matters into his own hands. Both examples demonstrate the simplistic handling of complex technology and contribute to workflow-related challenges. In both instances, the physicians later explained that they never train the system for the better if they do not use it. However, despite a slowness of the technology or a lower recognition rate than previously experienced, physicians keep using ASR when conducting EHRs. They have the opportunity to type in the record using a keyboard. Still, even though there is a sense of urgency and the next patient is waiting in the hallway, they remain until the algorithm is up to speed again before filling out records. Since physicians are free to use another medium, it does not seem to be a rule to use ASR, but an implicit and normative structure based on a goal of training the account which supports the illustrated activities such as getting coffee or writing down notes manually.

An interesting issue considers the measurement of an accounts' recognition rate and how it is visualized in graphs that do not necessarily reflect the physicians' use of the technology. During the first part of the fieldwork, the supplier provided a dataset containing variables measured on a monthly timeline over the past four years, illustrating "Dictations overturn-around time limit," "number of document lines," "unknown word rate," to name a few. The dataset is used to display distinctive recognition rates and use of time when producing EHRs. Once the physicians experience changes made when using the technology (which could be the abovementioned challenges in recognizing a word that

has been recognized so far or time delay related to the speed of word display), they are tempted to promptly criticize the technology for failing to live up to proclaimed expectations and listing possible obstacles to blame for the changes. Ongoing complications and changes would hypothetically affect measurable variables such as "unknown word rate" or "Dictations overturn-around time limit" and locate these obstacles or reason the physicians' indications. However, the dataset does not express the output measurements, and the supplier of the algorithm does not hold the values that result in an outcome of the dataset. Though the technology was implemented several years ago, the data extraction of its use presents itself as a somewhat black box containing variables.

Physicians form normative indications of possible obstacles created by sayings when using ASR. The missing explanation of why the technology fails to live up to an expected higher recognition rate leads to a loop of workarounds when using or failing to use ASR. Rather than zooming in on the dataset and attempt to make aspects more viewable than from a distance, another essential aspect is zoomed in on; the technology used in practice and its performative role. The use of technology when conducting EHRs is thus part of 'doing' ASR.

All the shadowed physicians are carriers of the practice of doing ASR and have different habits or workarounds regarding how they interact with computers, patients, and medical secretaries. An example is the abovementioned 'doing' of ASR, where a change of the technology leads to workarounds and to do another task than to see patients or conduct records. Further, to 'do' ASR is in this example related to another site than the orthopedic department and, more specifically, the ambulatory, since the use of the technology depends on external factors supplied by elsewhere.

## 6. Future Plans

So far, the study opens a limb to several practices concerning the use of ASR and further a possible understanding of how to 'do' ASR leads to a broader social order of healthcare professionals, suppliers, entities, and relations across these. This study will continue to analyze how these entities and further how the technology and the social are ordering practices and if these continue at other sites than the orthopedic departments or even hospitals. This thesis aims to identify practices associated with ASR and to assess the degree to which these appear or change while using the technology. The study also examines how workflows vary due to the complexity of the algorithm supporting ASR, as the newer version is built on a deep learning algorithm, and the previous is not. Forthcoming and more mature ASR algorithms are made based on the current designs and will continue to carry out practices found through this study. The results will contribute to the illumination of significant implications for future healthcare practices.

## 7. References

[1]   Alexandre Alapetite, Henning B. Andersen, and Morten Hertzum, 2009. "Acceptance of speech recognition by physicians: A survey of expectations, experiences, and social influence." International journal of human-computer studies 67, no. 1: 36-49.
[2]   Jens Edlef Møller and Henrik Vosegaard, 2008. "Experiences with electronic health records." IT professional 10.2: 19-23.
[3]   Yaron D. Derman, Tamara Arenovich, and John Strauss, 2010. "Speech recognition software and electronic psychiatric progress notes: physicians' ratings and preferences." BMC medical informatics and decision making 10.1: 1-7.
[4]   Tobias Hodgson, Farah Magrabi, and Enrico Coiera, 2017. "Efficiency and safety of speech recognition for documentation in the electronic health record." Journal of the American Medical Informatics Association 24.6: 1127-1133.
[5]   Frankie James, Jennifer Lai, Bernhard Suhm, Bruce Balentine, John Makhoul, Clifford Nass, and Ben Shneiderman, 2002. "Getting real about speech: overdue or overhyped?." In CHI'02 Extended Abstracts on Human Factors in Computing Systems, pp. 708-709.
[6]   Mike Wald and Keith Bain, 2007. "Enhancing the usability of real-time speech recognition captioning through personalised displays and real-time multiple speaker editing and annotation."

In International conference on universal access in human-computer interaction, pp. 446-452. Springer, Berlin, Heidelberg.

[7]   Jennifer Lai, Clare-Marie Karat and Nicole Yankelovich. 2007 "Conversational speech interfaces and technologies." The human-computer interaction handbook. CRC Press. 407-418.

[8]   Saswati Debnath and Pinki Roy, 2019 "Study of speech enabled healthcare technology." International Journal of Medical Engineering and Informatics 11, no. 1: 71-85.

[9]   Alexandre Alapetite, 2008. "Impact of noise and other factors on speech recognition in anaesthesia." International journal of medical informatics 77.1: 68-77.

[10]  Andreas S. Kirkedal, 2016. Danish Stød and Automatic Speech Recognition. Ph.D. dissertation, Frederiksberg: Copenhagen Business School (CBS).

[11]  Adam Burke, 2019. "Occluded algorithms." Big Data & Society 6.2: 2053951719858743.

[12]  Suzanne L. Thomas, Dawn Nafus, and Jamie Sherman, 2018. "Algorithms as fetish: Faith and possibility in algorithmic work." Big Data & Society 5.1: 2053951717751552.

[13]  Paul Dourish, 2016. "Algorithms and their others: Algorithmic culture in context." Big Data & Society 3.2: 2053951716665128.

[14]  Marina Kobayashi, Susan R. Fussell, Yan Xiao, and F. Jacob Seagull, 2005. "Work coordination, workflow, and workarounds in a medical context." In CHI'05 extended abstracts on Human factors in computing systems, pp. 1561-1564.

[15]  Deborah S. Debono, David Greenfield, Joanne F. Travaglia, Janet C. Long, Deborah Black, Julie Johnson, and Jeffrey Braithwaite, 2013. "Nurses' workarounds in acute healthcare settings: a scoping review." BMC health services research 13, no. 1: 175.

[16]  Karin Knorr-Cetina, Eike von Savigny, and Theodore R. Schatzki, 2001. eds. The practice turn in contemporary theory. Routledge.

[17]  Davide Nicolini, 2012. Practice theory, work, and organization: An introduction. OUP Oxford, 2012.

[18]  Theodore Schatzki, 2002. The site of the social: A philosophical exploration of the constitution of social life and change. University Park, PA: Pennsylvania State University Press.

[19]  Davide Nicolini, 2009. "Zooming in and out: Studying practices by switching theoretical lenses and trailing connections." Organization studies 30.12: 1391-1418.

[20]  Steinar Kvale and Svend Brinkmann, 2009. Interview: introduktion til et håndværk. Hans Reitzels Forlag.

[21]  Bente Halkier, 2016. 201. Fokusgrupper (3. udg. ed.). Frederiksberg: Samfundslitteratur.

[22]  Seonaidh McDonald, 2005. "Studying actions in context: a qualitative shadowing method for organizational research." Qualitative research 5.4: 455-473.

[23]  Johannes Fabian, 2004. "On Recognizing Things. The "Ethnic Artefact" and the "Ethnographic Object"." L'Homme. Revue française d'anthropologie 170: 47-60.