Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 67 (2009) 36–49
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Acceptance of speech recognition by physicians: A survey of expectations, experiences, and social influence

Alexandre Alapetite[a,b,*], Henning Boje Andersen[a], Morten Hertzum[b]

[a]*Systems Analysis Department, Risø National Laboratory, Technical University of Denmark, DK-4000 Roskilde, Denmark*
[b]*Computer Science, Roskilde University, Universitetsvej 1, P.O. Box 260, DK-4000 Roskilde, Denmark*

Received 30 October 2007; received in revised form 20 June 2008; accepted 15 August 2008
Communicated by C.M. Karat
Available online 22 August 2008

Abstract

The present study has surveyed physician views and attitudes before and after the introduction of speech technology as a front end to an electronic medical record. At the hospital where the survey was made, speech technology recently (2006–2007) replaced traditional dictation and subsequent secretarial transcription for all physicians in clinical departments. The aim of the survey was (i) to identify how attitudes and perceptions among physicians affected the acceptance and success of the speech-recognition system and the new work procedures associated with it; and (ii) to assess the degree to which physicians' attitudes and expectations to the use of speech technology changed after actually using it. The survey was based on two questionnaires—one administered when the physicians were about to begin training with the speech-recognition system and another, asking similar questions, when they had had some experience with the system. The survey data were supplemented with performance data from the speech-recognition system. The results show that the surveyed physicians tended to report a more negative view of the system after having used it for some months than before. When judging the system retrospectively, physicians are approximately evenly divided between those who think it was a good idea to introduce speech recognition (33%), those who think it was not (31%) and those who are neutral (36%). In particular, the physicians felt that they spent much more time producing medical records than before, including time correcting the speech recognition, and that the overall quality of records had declined. Nevertheless, workflow improvements and the possibility to access the records immediately after dictation were almost unanimously appreciated. Physicians' affinity with the system seems to be quite dependent on their perception of the associated new work procedures.

Keywords: Speech recognition; Technology acceptance; Electronic medical records

1. Introduction

Speech recognition has been refined and become more robust in recent years (Lai et al., 2008). The gradual maturation of the technology has been accompanied by adoptions of the technology in the medical domain, where it is used to enter comments into the electronic medical record (EMR), thus replacing the standard way of entering notes by physician dictation and subsequent transcription by medical secretaries or a dedicated service (Zafar et al., 1999; Al-Aynati and Chorneyko, 2003). At the same time as the technology has matured, speech recognition has been developed and implemented for languages spoken by much "smaller" populations, such as Danish (5.4 million speakers).

Vejle and Give Hospital, Denmark, has been one of the first hospitals to introduce speech recognition for all major specialties and departments. Having run a successful project on speech recognition in its radiology department since 2000, this regional hospital (349 beds, and 217 000 outpatients in 2006) began to implement plans for having all physicians in clinical departments use speech recognition to input physician notes and instructions into the

*Corresponding author at: Systems Analysis Department, Risø National Laboratory, Technical University of Denmark, DK-4000 Roskilde, Denmark.
E-mail address: alexandre@alapetite.net (A. Alapetite).

1071-5819/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2008.08.004

EMR. The speech-recognition system—software based on Philips Speech Magic, adapted to Danish and deployed by Max Manus A/S—was rolled out in all clinical departments in 2005–2006, and has about 240 physician users as of 2007.

The main purpose of introducing speech recognition across all departments was to ensure a quicker completion of medical record entry and to achieve a higher quality of patient records. The old transcription service was known to sometimes produce backlogs of dictation tapes waiting to be transcribed by medical secretaries, or transcriptions waiting to be checked and approved by physicians. Additionally, an expected consequence was to allow secretaries, who would no longer need to spend time on transcriptions, to take over other duties. It was hoped that the quality of medical records would be enhanced, since physicians would now be going to check and revise their written (speech recognized) record immediately while their intentions were still fresh in memory. While little is known so far about the impacts of speech recognition on the various stages of the writing process and on the quality of outcome (Honeycutt, 2003), the above-mentioned goals fully match criteria such as those reported by Mönnich and Wetter (2000).

The present study had two related objectives: First, to identify physicians' attitudes and expectations about speech recognition that might explain their subsequent level of satisfaction with actual use of the technology. Second, to assess possible changes between prior expectations to and subsequent experience with the technology as a replacement for the traditional mode of producing medical records.

2. Related work

Work about the acceptance of speech recognition falls into two main areas: speech recognition and technology acceptance. Studies of speech recognition have predominantly been devoted to recognition of spoken English. However, recognition rates of systems that recognize English are not necessarily transferable to a speech-recognition system for Danish.

2.1. Speech recognition

For free-text dictation, speech recognition combines some characteristics of traditional dictation and of word processing (Leijten and Van Waes, 2005): on the one hand, quick and easy use of speech, and on the other, instantaneous graphical feedback and the possibility of jumping back and forth in the text. At the same time, speech recognition has its own advantages and drawbacks.

For transcription of free text, state-of-the-art systems correctly recognize 72–98% of the spoken words according to recent research (Zafar et al., 1999, 2004; Devine et al., 2000; Jungk et al., 2000; Ramaswamy et al., 2000; Kanal et al., 2001; Sears et al., 2001; Al-Aynati and Chorneyko, 2003; Alapetite, 2008), while commercially reported recognition rates are generally above 95%. Several factors contribute to the differences in recognition rates across studies:

- Vocabulary affects speech recognition through its size and domain coverage. Large vocabularies with good domain coverage are attractive, simply because they enable recognition of more words. Conversely, the acoustic distinctiveness of words is larger in small vocabularies, increasing the likelihood of correct recognition. Small vocabularies are, however, mostly relevant for voice navigation. State-of-the-art systems for text transcription have vocabularies comprising tens of thousands of words and optional, add-on vocabularies for specific domains such as the medical domain.
- Speakers influence speech recognition by the clarity and consistency of pronunciation and the degree of fit between their pronunciation and the acoustic model of the system. Speaker-dependent systems achieve higher recognition rates than speaker-independent systems but require one or more training sessions—based on which the system adapts its acoustic model to the speaker—and may be more sensitive to variations of the background noise, microphone, and voice (e.g., due to a cold). Even after training, atypical speakers, including non-natives (Coniam, 1999) as well as children and elderly (Wilpon and Jacobsen, 1996), experience lower recognition rates.
- Noise affects speech recognition in two ways: (a) it distorts the speech signal, making it more difficult to discern the spoken words. (b) In the presence of noise, people alter their voice in an attempt to counter the distortion of the speech signal (the Lombard effect; Lombard, 1911). Ambient noises, such as those heard in hospital wards or emergency rooms, do not significantly affect average speech-recognition rates (Zafar et al., 1999; Alapetite, 2008). However, in spite of numerous noise-cancellation techniques, loud noise, and even moderate levels of noise, may considerably degrade the performance of speech-recognition systems (Gong, 1995; Barker et al., 2005).
- All speech-recognition systems are based on principles of statistical pattern matching (Young, 1996). However, in spite of this commonality, individual systems differ in their parameterization of the speech signal, the acoustic model of each phoneme, and the language model used in predicting the words most likely to follow the preceding words. Thus, different systems make different recognition errors, even when they achieve similar recognition rates. This difference can be used to improve recognition rates by fusing the outputs of multiple systems (Fiscus, 1997; Alapetite, 2008).

Studies of text transcription show that it takes more time for a person to produce a text by voice input followed by correction of the recognition errors than by dictation