

# The Unintended Irregularities of Automatic Speech Recognition

Silja Vase

Department of Communication/ University of Copenhagen  
Karen Blixens Plads 8, 2300 Copenhagen S, Denmark  
Siljavase@hum.ku.dk

## Extended Abstract

The present study examines the emerging role of Automatic Speech Recognition (ASR) and the unintended irregularities that arise when the algorithm is configured in healthcare practices. Once you consider health information technology, the mind often drifts towards new digital devices or tools that are implemented to improve quality or create efficiency. Scholars within Human-Computer Interaction (HCI) reflect widely upon how digital technology has become a vital part of healthcare exploring the Internet of Things (IoT) [1], visually directed applications [2] and usability frameworks [3] to name a few perspectives. However, this study focusses on ASR, an algorithm that is used by physicians to conduct Electronic Health Records (EHRs), which allows the patients to access their records in real-time [4]. ASR substitutes dictation and typing when physicians produce these medical records.

The central objective is to understand how physicians navigate in relationship to ASR and how the algorithm affects practices. The study is based on ethnographically collected data at a Danish hospital in Jutland, where the American developed Nuance SpeechMagic algorithm was implemented approximately ten years ago and has become a mandatory medium for physicians when they conduct EHRs. The study is based on the data collected at a multi-sited study that starts at an orthopedic department and follows the algorithm around through several departments and ends at the supplier. In short, ASR enables natural language to be translated into text, which subsequently can be edited by physicians to train the algorithm for a better translation in the future [5]. The specific algorithm studied is based on a linear algorithm, trained by corrections made by physicians who represent diverse practices in various fields. The algorithm is thus expected to cover the disciplines and semantic representations within these when documenting EHRs. Further, ASR is expected to comprehend interferences or malfunctions of other technologies, e.g., phone calls, alerts, 'frozen' accounts or screens, uneven connection, timeworn microphones, etc.

The study shows how the performance measurements visualized by ASR do not reflect the physicians' use of it, which is demonstrated as practice-related challenges. The challenges emerge from 1) lack of technical literacy, 2) interference made by the surroundings, or 3) technical malfunction. This study shows how the present algorithm does not meet expectations made by physicians towards the technology and how it further fails to deliver a reduction of time spent on producing medical records. What is more, individual physicians need to spend more time training the algorithm than others, as their speech does not fit into predetermined gender-related categories, which in turn complicates their time spent on patients. The findings further meet concerns stated by previous voices in the field, such as how non-native speakers can experience a lower recognition of their voices [6] and how noise also distorts the recognition ability [7].

Upcoming deep learning ASR algorithms are based on the current design and will continue to produce unintended irregularities. As this study is based on a linear ASR algorithm, future studies should thus focus on the cognitive aspects to research significant implications for future healthcare practices.

## References

- [1] A. K. M. Majumder, Y. A. ElSaadany, R. Young and D. R. Ucci, "An energy efficient wearable smart IoT system to predict cardiac arrest" in *Advances in Human-Computer Interaction*, vol 2019, Hindawi, 2019.
- [2] B. Reiner, "Visually directed human-computer interaction for medical applications." U.S. Patent No. 9,841,811. 2017.
- [3] R. J. Holden, S. Volda, A. Savoy, J. F. Jones, and A. Kulanthaivel, "Human factors engineering and human-computer interaction: supporting user performance and experience" in *Clinical informatics study guide*, pp. 287-307. Springer, Cham, 2016.

- [4] T. Hodgson, F. Magrabi and E. Coiera, "Efficiency and safety of speech recognition for documentation in the electronic health record" in *Journal of the American Medical Informatics Association*, vol. 24, no.6, pp. 1127-1133, 2017.
- [5] A. S. Kirkedal, "Danish Stød and Automatic Speech Recognition" Ph.D. dissertation, Copenhagen Business School, 2016.
- [6] D. Coniam, "Voice recognition software accuracy with second language speakers of English" in *System*, vol. 27, no. 1, pp. 49-64, 1999.
- [7] A. Alapetite, H. B. Andersen, and M. Hertzum "Acceptance of speech recognition by physicians: A survey of expectations, experiences, and social influence" in *International journal of human-computer studies*, vol. 67, no. 1, pp. 36-49, 2009.