Editorial: Towards Omnipresent and Smart Speech Assistants

Ingo Siegert [1]*, Stefan Hillmann [2], Benjamin Weiss [2], Jessica M. Szczuka [3] and Alexey Karpov [4]

[1] Mobile Dialog Systems, Institute for Information Technology and Communications, Otto von Guericke University Magdeburg, Magdeburg, Germany, [2] Quality and Usability Lab, Technische Universität Berlin, Berlin, Germany, [3] Social Psychology: Media and Communication, University of Duisburg-Essen, Duisburg, Germany, [4] St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, Russia

Editorial on the Research Topic

Towards Omnipresent and Smart Speech Assistants

1. INTRODUCTION

The functionality of digital voice assistant systems has been constantly increasing during the last decade and a lot of commercial systems are available. Driven by their ease of use, the attractiveness of such devices is constantly growing, and they allow conducting online searches and orders as well as smart home services by simply calling up the device (de Barcelos Silva et al., 2020; Dutsinma et al., 2022).

However, the implications of voice-based interaction are not always clear to the user, ranging from its functionality to the impact of speech as a social cue for resulting psychological effects. In the future, however, they should not only process simple commands, but also enable a natural and smooth interaction and be omnipresent. In addition to an improved speech recognition, this will require enhanced speech understanding and more intelligent dialog guidance.

While state-of-the art systems are mainly conceptualized for young adults and middle-aged people, future systems should adapt to the user in order to meet the needs of different (vulnerable) user groups, ranging from young children to the elderly. This will be accompanied by efforts to make systems more understandable and users more sophisticated. Consequently, legal aspects resulting from the spread of voice assistants and the stricter data protection regulations are important.

The goal of this Research Topic was to present the latest advances—both from academia and industry—in the area of voice assistants. It was aimed at collecting research contributions from the disciplines of human-computer interaction, artificial intelligence, and human factors in order to promote interdisciplinary collaborations and cross-fertilization of ideas. More specifically, we were interested in exploring the current landscape and future directions for the emerging topic of voice assistants. The Research Topic covers 11 articles from 34 different authors from different research fields, including linguistics, psychology, usability/user experience studies as well as the technical perspective. One apparent focus of this Research Topic was on analyzing and assessing user experience. Both, different user groups and situations are taken into account. However, we hope to see the aforementioned perspective on more sophisticated dialogs represented in the near future.

2. CONTRIBUTIONS

Cao et al. investigate how mind-based anthropomorphism influences users exploratory usage of intelligent personal assistants (IPA). The article describes a study collecting more than 500 valid answered questionnaires, and the results on the influence of cognitive and affective anthropomorphism on IPA self-efficacy and the user's social connection to the IPA.

Carolus et al. show in an online laboratory experiment that participants have empathy with a smart speaker, when watching videos of a user interacting with such a device. This claims a rather universal effect, as the results are independent of the participants' gender or usage experience, and thus expands the current body of empirical results around the Media Equation (Reeves and Nass, 1996).

Cohn et al. investigate users speech rate adjustments during conversations with an Amazon Alexa social bot in at-home and in-lab settings, considering automatic speech recognition (ASR) comprehension errors. It is found that users used a slower speech rate when talking to the bot, which is even more slowed down in the in-lab setting (relative to at-home).

Cohn and Zellou present the results of a study on differences in speech adaptations (e.g., speech rate, f0 mean, and f0 variation) during pre-scripted spoken interactions with a voice-AI assistant and a human interlocutor. The authors measured a decreased speech rate, higher average fundamental frequency (f0), and greater f0 variation for the device directed speech.

Frommherz and Zarcone collected ecologically-valid German dialog data via a crowdsourcing approach in the Wizard-of-Oz (WOZ) setting. Compared with the MultiWOZ dataset, their method for data collection has led to considerably less scripting and priming in the collected dialog data.

Hirsch presents a local and low-cost, low-energy voice assistant solution including a keyword recognition algorithm and a further client system without the need of an external power supply. This is the most relevant applied work, of a privacy-ensuring home speech assistant, among all the articles.

Mavrina et al. describe a longitudinal field study on communication breakdowns between family members and a voice assistant. Their article provides qualitative analysis of particularly interesting breakdown cases, as well as statistical analysis combining empirical and conversational data collected with children and adults during 5 weeks of free interaction with a voice assistant device.

Schlomann et al. present their opinion regarding elderly with and without cognitive disabilities. Their main argument is to raise the potential of speech assistants for elderly users by participatory design methods and verify the approaches by field studies.

Schreibelmayr and Mara conducted a randomized laboratory experiment on synthetic voices with 165 participants to explore what level of human-like realism human-interactors prefer, whether the participants evaluations vary across different domains of application, and if the listener's personality has an impact on the ratings.

Wienrich and Carolus have developed an instrument called "conversational agent literacy scale" (CALS), to measure conceptualizations and competencies about conversational agents in human users. This scale consists of five sub-scales and is based on a study with 170 participants.

Wienrich et al. found in a laboratory study that a voice assistant designed as a "specialist" is rated as more trustworthy by the users than a "generalist" in the health domain. By providing both, a theoretical line of reasoning and empirical data, the study lays the pathway for further studies on the users perspective on trustworthiness in voice-based systems.

3. CONCLUSION

In conclusion, this Research Topic comprises interdisciplinary contributions and gives some examples of both theoretical and practical implications for smart voice/speech assistants. Topics reach from laboratory studies on empathy or speaking behavior adjustments over field studies on communication breakdowns, to the description of a local client voice assistant system. It therefore reflects the diversity of this strongly developing field of research. However, the contributions also highlight unresolved questions in current research, e.g., pitfalls due to design and field study issues or a lack of studies regarding trust or acceptance.

We are aware that there is a plethora of further aspects that need to be addressed to complete, in the best sense, the aim of a human-like interaction with voice assistants for all kind of humans. The articles of this Research Topic paving the way to an understanding of the role of voice assistants and thus, in the future, voice assistants can be an integral part of our daily life in terms of a true intelligent assistant.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

de Barcelos Silva, A., Gomes, M. M., da Costa, C. A., da Rosa Righi, R., Barbosa, J. L. V., Pessin, G., et al. (2020). Intelligent personal assistants: a systematic literature review. *Expert Syst. Appl.* 147, 113193. doi: 10.1016/j.eswa.2020.11 3193

Dutsinma, F. L. I., Pa, D., Funilkul, S., and Chan, J. H. (2022). A systematic review of voice assistant usability: An ISO 9241-11 approach. *SN Comput. Sci.* 3, 267. doi: 10.1007/s42979-022-01 172-3

Reeves, B., and Nass, C., editors (1996). *The Media Equation: How People Treat Computers, Television, and New*

Media like Real People. Cambridge, UK: University Press Cambridge.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.