
Introduction to the Workshop

Astrid Carolus¹, Carolin Wienrich², and Ingo Siegert³

¹Julius-Maximilians-University, Wuerzburg, astrid.carolus@uni-wuerzburg.de

²Julius-Maximilians-University, Wuerzburg, carolin.wienrich@uni-wuerzburg.de

³Otto-von-Guericke-University, Magdeburg, ingo.siegert@ovgu.de

The Workshop on **Establishing an interdisciplinary perspective on speech-based technology** has the aim to collect and connect theoretical and methodological approaches from multiple disciplines to widen the perspective on speech-based technology.

Motivation and Goals of the Workshop

Speech-based technology has evolved dramatically over the last few years by technical improvements from data science/ engineering science, especially in far-field acoustic speech enhancement, together with a vast performance increase in automatic speech recognition and speech understanding, on the one side. Particularly noteworthy is the field of Affective Computing – aiming to develop technical systems able to receive, understand, and (adequately) react to affective signals [2]. As well as the recent improvements in artificial intelligence (i.e., Deep, Ensemble, Active Learning, and Fusion) – allowing to improve and adapt the system constantly and directly to a user [6, 3]. Furthermore, ubiquitous technological advances provide the technology always at the user’s location [4]. Thus, technological developments have changed the human-technique interaction fundamentally, resulting in vital consequences for the human users, on the other side. Speech-based technology steps over from pure command receivers to a kind of companion system that incorporates human needs, desires, and abilities [5]. Therefore, interdisciplinary research activities need to be intensified integrating the various scientific perspectives beyond a mere technological focus.

With its focus on the individual, the psychological perspective indicates the increasing similarities between the operation of speech-based technology and human-human interaction. The human user experiences some kind of conversation with the device or with an application, which seems to listen and answer. Conducting a conversation, however, has been a uniquely human ability until recently [1]. Adopting these allegedly human qualities results in the usage of technology to fulfill criteria, which have been exclusive for human-human interactions [7]. Consequently, from the human user’s perspective, the line between using a technological device or application and interacting with an (at least) humanlike counterpart has begun to blur. Consequently, the “human” in human-computer interaction has become increasingly relevant, enhancing the relevance of others than technology-oriented disciplines (e.g., data science, engineering

science, media science, HCI). If “computers” transfer into humanlike counterparts, disciplines, which have established a treasure of knowledge on humans and their “functional principles”, will be valuable sources of knowledge. They already know how humans think, feel and behave. They know about the way they learn, how they live together, communicate, collaborate. Social sciences (e.g., sociology, educational and political science) or humanities (e.g., psychology, philosophy, history, art) , and medical and health sciences could draw on a large body of knowledge, theoretical concepts or methodological approaches, which research on speech-based technology can benefit from. However, literature reviews reveal a narrow field of contributing disciplines, with most studies originating from standard natural or technical sciences. In contrast to other rather new digital technologies such as social robots or virtual assistants, research on speech-based technology is only at the beginning of the process to become an established area of research- although, for example, voice assistants can already be found in millions of households.

The first **workshop, “AI Debate”**, aims to widen the perspective on human-technology interaction by bringing together researchers from different scientific areas rarely represented in the field. The contribution focuses on three main aims:

1. connecting researchers from different backgrounds
2. collecting theoretical and methodological approaches from the range of scientific disciplines
3. analyzing possible conceptual and methodological overlaps and cross connections

The present contribution summarizes the input comments given by researchers of different areas participating in the workshop. The comments answer five questions.

1. What do I know already? By answering this question, researchers report theories and concepts on speech-based technology from their perspective.
2. How do I study the phenomenon? Answers to this question are current methodological approaches to studying speech-based technology.
3. What would I like to know? Here, researchers formulate open questions and future directions to learn about speech-based technology from other research areas.
4. What do I want to learn from different disciplines? Researchers reify specific theoretical or methodological approaches from other disciplines to gain deeper insights into speech-based technology by answering this question.
5. What do I want to teach other disciplines? These answers are specific theoretical or methodological approaches from their field they wish to share to gain deeper insights into speech-based technology.

The Organizers

Astrid Carolus is an assistant professor for media psychology at the University of Würzburg. In her research, she focuses on humans interacting with digital technology and with media content. From a psychological perspective, she analyzes the underlying motives of media usage (as well as the refusal of usage) and its effects. Her research foci are on human users interacting with digital devices and AI-enabled technology, which

she conceptualizes as psychologically relevant entities; on health-related effects of social media communication; and on teachers' digital literacy.

Carolyn Wienrich is a professor for Human-Technique-Systems at the University of Würzburg and co-leader of the XR HUB Würzburg. Her research interests focus on interaction paradigms between humans and digital entities and change experiences during and after digital interventions. Her team explores antecedents, potentials, and risks of digital interactions and experiences since digital entities and digital interventions accompany humans in many contexts. Participative and human-centered research, theoretical concepts, and multi-methods stemming from psychology and computer science define her qualification in human-computer interaction.

Ingo Siegert is assistant professor for Mobile Dialog Systems at the Otto von Guericke University Magdeburg. His research interests are on signal-based analyses and interdisciplinary investigations of human-computer interaction in terms of addressee detection, perceived charisma as well as the utilization of further interaction patterns, such as filled pauses or discourse particles within the scope of voice assistants. Ingo Siegert has published 100+ peer reviewed papers and articles on several conferences and various journals and is co-organizer of various workshops and conferences.

References

- [1] S. Pinker. *The language instinct. How the Mind Creates Language*. William Morrow & Co, USA., 1994.
- [2] B. Schuller et al. "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge". In: *Speech Commun* 53 (9-10 Nov. 2011), pp. 1062–1087.
- [3] M. Glodek et al. "Fusion paradigms in cognitive technical systems for human–computer interaction". In: *Neurocomputing* 161 (2015), pp. 17–37.
- [4] S. Biundo and A. Wendemuth. "Companion-Technology for Cognitive Technical Systems". In: *KI - Künstliche Intelligenz* 30.1 (Feb. 2016), pp. 71–75.
- [5] T. Gossen et al. "Modeling aspects in human-computer interaction - adaptivity, user characteristics and evaluation". In: *Companion technology: a paradigm shift in human-technology interaction*. Cham: Springer International Publishing, 2017, pp. 57–78.
- [6] F. Schwenker et al. "Multimodal affect recognition in the context of human-computer interaction for companion-systems". In: *Companion technology: a paradigm shift in human-technology interaction*. Cham: Springer International Publishing, 2017, pp. 378–408.
- [7] A. Carolus et al. "Impertinent mobiles - Effects of politeness and impoliteness in human-smartphone interaction". In: *Computers in Human Behavior* 93 (2019), pp. 290–300.