

Evaluating conversational systems: Concepts, theories, and methods

Benjamin Weiß (Technische Universität Berlin)

The current trend in interacting with computers is voice—albeit the dominant paradigm is the one-shot interaction. However, when engaging in a conversations with more than just one turn, evaluation becomes tricky. After introducing a functional architecture of conversational systems, I will present theories on human interaction and derive potential evaluation concepts to support a systematic approach to identifying and subsequently evaluating the desired quality aspect(s) of an interactive system. Empirical methods comprise listening tests (e.g. for Text2Speech quality), laboratory experiments, and field-like methods. The potential application of established approaches from engineering (aggregating log-data over time, PARADISE framework), social psychology (social facilitation-inhibition, attributions), and linguistics (Bernsen’s extension of Grice’s Maxims) will be discussed as well.