Usability and Performance Evaluation for Advanced Spoken Dialogue Systems

Kristiina Jokinen
Department of Computer Sciences
University of Tampere / University of Tartu
kristiina.jokinen@helsinki.fi

Michael McTear
School of Computing and Mathematics
University of Ulster at Jordanstown
MF.McTear@ulster.ac.uk

Abstract

The past decade has seen a rapid emergence of dialogue systems that support robust and efficient interaction in spoken natural language. The technology has become mature enough for speech-based interactive applications to be built for practical purposes, and the results are also being applied to novel fields such as situated cognition and robots, embodied conversational agents, meeting assistants, etc. There has also been a growth of research focusing on advanced spoken dialogue systems that aim to increase the system's communicative competence by including aspects of interaction that go beyond the basic techniques of interaction management. Such advanced aspects include e.g. disfluencies, turn-taking, speaker intentions, emotions, multimodality, and adaptation in context.

One of the motivations for furthering the system's interaction capabilities is to improve the system's naturalness and usability in practical applications. However, there are several issues that need to be considered, and requirements and evaluation metrics seem to differ for academic and industrial perspectives. For instance, over the past decade, the research community has focussed on various data-driven methods in contrast to the hand-crafted rules that are used predominantly in commercial systems. The main arguments in support of the data-driven approach concern robust understanding and the assumption that it is ultimately more portable and less labour intensive than hand-crafting, whereas approaches for practical systems usually emphasise rather simple albeit robust solutions, and the importance of adhering to the requirements, needs and preferences of real users.

Research and development presuppose well-defined criteria according to which interactive systems can be evaluated in terms of usability and naturalness. Objective and subjective criteria have been identified and enumerated, and although no consensus has been reached on the general practices, criteria, or metrics of evaluation, it is widely agreed that rigorous evaluation methodologies should be consolidated as part of the research activities, especially when dealing with complex issues as such adaptation, user requirements, and best-practice applications.

This tutorial will focus on methods, problems and challenges in the evaluation of advanced spoken dialogue systems. It is grounded in research that combines various speech and language technology components into an integrated system, and surveys the issues related to the design, evaluation and comparison of such systems. A number of different approaches to robust and efficient interaction management will be reviewed, together with various performance and user evaluation methods in academic and industrial environments. A closer look will be taken at different metrics and usability criteria, as well as automatic design and evaluation methods. Practical requirements for dialogue systems, such as robustness, scalability and portability will also be discussed and exemplified from the point of view of performance evaluation and usability. Special attention will be paid to user evaluation, and to the user's expectations and experience of the system.

References

Jokinen, Kristiina. forthcoming. *Constructive Dialogue Management – Speech Interaction and Rational Agents*. John Wiley & Sons.

Michael McTear. 2004. *Spoken dialogue technology: toward the conversational user interface* Springer Verlag.