

Preface

Spoken Dialog Systems (SDS) became popular in the past years for telephone-based services such as database queries (e.g., bus information), banking, or troubleshooting (e.g., Internet equipment). Due to the recent improvements in automatic speech recognition and natural language understanding, these systems have reached a degree of complexity which makes their design a challenge for developers. Thus, it has become increasingly important to evaluate the systems to ensure that performance criteria are met and users are satisfied with the service. While methods exist to assess system quality with users, these are often used less than required, as they are demanding in terms of time, money, and expertise.

Automatic evaluation has been proposed as a way to reduce the costs of user testing. This book examines how such methods can be applied to SDSs, and how they can be integrated in a system design process. More specifically, the discussed methods are based on user models, meaning that the involved algorithms reflect certain aspects of the behavior expected from the prospective users of the system. On the one hand, a model of the users' interaction behavior can be used to simulate dialogs with the system, and by this generate performance estimates and detect dialog design problems. On the other hand, a model of the users' judgment behavior allows predicting user ratings about the perceived quality of the dialogs. Although both types of user models can hardly be defined completely, it is assumed that they can, nevertheless, be of value during system design. The aim of the book is, thus, to analyze the performance of these models with respect to different criteria which might be employed in this process.

I wrote this book during my past four years at the Quality and Usability (QU) Lab at TU Berlin. As part of Telekom Innovation Laboratories (T-Labs), the QU-Lab offered me insights into many different directions research can take. I am grateful for having the opportunity to work here, and I acknowledge the positive impact a number of people I met during this time had on my work in one way or the other.

Generally, I have experienced most of the colleagues at T-Labs, as well as in the international dialog systems research community, to be open for any kind of discussions, which helped shaping ideas as well as forgetting them.

In particular, I would like to thank some colleagues for more specific contributions:

- The MeMo workbench described in [Chap. 2](#) was planned and implemented during the MeMo project funded by T-Labs, which was conducted in collaboration with DFKI (Anthony Jameson, Michael Kruppa, Angela Mahr), DAI-Labor (Thimios Dimopoulos, Marco Kunze, Maximilian Kern, Andreas Rieger, Michael Quade, Mathias Runge, Carsten Wirth) and Fraunhofer FIT (Britta Hofmann, Dorothea Kugelmeier, Mark Hümmer). I am particularly grateful that Michael Quade implemented some functions needed for the evaluation of the INSPIRE simulation in [Chap. 3](#).
- Florian Gödde and Mark Hümmer conducted the user test described in [Chap. 3](#). Florian also made transcriptions and some annotations of this corpus and helped conducting the experiment described in [Chap. 5](#). Stefan Schmidt co-annotated the interaction problems discussed in [Chap. 4](#) and proof-read this chapter.
- Felix Hartard did his Diploma thesis on predicting user judgments with Hidden Markov Models, and helped in the data collection described in [Chap. 5](#) as Wizard-of-Oz. For the same experiment, Ulrike Stiefelbogen lent her voice to the dialog system mock up. I am also grateful for the advice regarding HMMs I got from Hamed Ketabdar.
- Babette Wiezorek worked for me as a student worker for 2 years and made my life easier by taking over many annotations in the databases I used, as well as supporting other work which did not make its way into this book.
- Robert Schleicher often advised me in statistics and other methodological questions.
- Benjamin Weiss read the entire manuscript from top to bottom, providing lots of comments, and feedback, which helped me improving it, without being demotivating.
- I greatly appreciate that Prof. Ramón López-Cózar kindly took over the co-assessment of my dissertation, on which this book is based, and provided many useful comments.
- I am particularly grateful to my dissertation supervisor, Prof. Sebastian Möller, head of the QU-Lab, for the great time I had in the past 4 years, and for the dedicated supervision of my work during that time. Sebastian had an open ear for all kinds of problems, and the many discussions we had greatly formed the content of this book.
- Finally, I would like to thank my friends for giving me true spare time, my parents for taking care for and teaching me to value a good education, and Ina for accompanying me in all I do.

The publication of this book was enabled by the smooth cooperation with Christoph Baumann at Springer, who kindly accepted the manuscript for publication and kept the organization simple for me. In addition, earlier versions of [Chaps. 2, 3](#) and [5](#) have been published previously, and I list the original sources

below. All the original publishers have been contacted, and their permission to reuse these articles including figures and tables is hereby acknowledged:

- Engelbrecht, K.-P., Kruppa, M., Möller, S., Quade, M., 2008. MeMo Workbench for Semi-Automated Usability Testing, in: Proc. of Interspeech 2008, Brisbane, Australia (text passages, partly modified, used in Chap. 2).
- Engelbrecht, K.-P., Quade, M., Möller, S., 2009. Analysis of a New Simulation Approach to Dialogue System Evaluation. *Speech Communication*, 51, pp. 1234–1252 (text passages, figures and tables used in the Introduction, Chap. 2 and Chap. 3).
- Engelbrecht, K.-P., Möller, S., 2010. Sequential Classifiers for the Prediction of User Judgments about Spoken Dialog Systems. *Speech Communication*, 52, pp. 816–833 (full article used in slightly modified version as Chap. 5, text passages used in the Introduction).



<http://www.springer.com/978-3-642-31590-9>

Estimating Spoken Dialog System Quality with User
Models

Engelbrecht, K.-P.

2013, XIV, 130 p., Hardcover

ISBN: 978-3-642-31590-9