

Book Reviews

Spoken Dialogue Technology: Toward the Conversational User Interface

Michael F. McTear
(University of Ulster)

London: Springer-Verlag, 2004,
x+432 pp; paperbound, ISBN
1-85233-672-2, \$59.95

Reviewed by
Johannes Pittermann
University of Ulm

What would you say if your refrigerator told you, “You’re having some friends round for hot chocolate later. Maybe you should order two cartons of milk”? Of course, in *Spoken Dialogue Technology*, Michael McTear will not give an answer to the question of whether talking to domestic appliances makes sense, but he indicates that even a normal household, for instance, may offer a wide field of application for spoken-language dialogue systems in the near future. Consequently his book primarily focuses on theory and practice of these systems.

Addressing undergraduate students as well as postgraduate researchers and practitioners in human-computer interfaces, the book is subdivided into three parts which meet the readers’ needs: “Background to Spoken Dialogue Technology” (Chapters 1–5), “Developing Spoken Dialogue Applications” (Chapters 6–11), and “Advanced Applications” (Chapters 12–14).

Chapter 1, “Talking with Computers: Fact or Fiction,” and Chapter 2, “Spoken Dialogue Applications: Research Directions and Commercial Deployment,” present recent products and aspects of dialogue technology as well as historical linguistic and artificial intelligence approaches to dialogue and simulated conversation. Aspects of present-day commercial use of spoken dialogue technology are also discussed. In Chapter 3, “Understanding Dialogue,” the term *dialogue* is defined, and four of its key characteristics—dialogue as discourse, dialogue as purposeful activity, dialogue as collaborative activity, and utterances in dialogue—and its structures and processes are described in detail. Chapter 4 gives an overview of the components of a spoken language dialogue system: speech recognition, language understanding, language generation, and text-to-speech synthesis. The central component (i.e., dialogue management) is specified in Chapter 5. Here, dialogue initiative (system initiative, user initiative, and mixed initiative), dialogue control (finite-state-based, frame-based, and agent-based control), and grounding (how to process the user’s input) are described. Furthermore, knowledge sources (dialogue history, task record, world knowledge model, domain model, generic model, and user model) and problems that arise when interacting with an external knowledge source are discussed.

The second part starts with dialogue engineering, which can be subdivided into analysis and specification of requirements, design, implementation, testing, and evaluation of a dialogue system. The use-case analysis includes user profile (type of user, language, user’s experience level, etc.) and usage profile (frequency of use, input/output device type, environment, etc.). The spoken-language requirements can

be analyzed with the help of existing corpora or simulations using the Wizard of Oz method. In the requirements specification, the developer defines what the system is intended to do. How these specifications are achieved is defined in the design phase. This comprehends the dialogue flow, prompts, grammars, interaction style, navigation, help, confirmation, etc. For the implementation, McTear describes the CSLU Toolkit and its Rapid Application Developer (RAD), VoiceXML platforms, and platforms for multimodal Web-based applications like Microsoft's SALT (Speech Application Language Tags), some of which are dealt with in the subsequent tutorial chapters. In the testing and evaluation sections, several test and evaluation methods are outlined. Whereas testing is necessary to determine whether the system conforms to the specifications, the evaluation phase comprises the analysis of user acceptance and the analysis of the system performance which can be accomplished, for example, by using the PARADISE tool.

The subsequent chapters of the second part consist of tutorials on how to develop and implement a dialogue system. Chapter 7, "Developing a Spoken Dialogue System Using the CSLU Toolkit," deals with the development of spoken dialogue systems with the help of RAD. This chapter starts with the basic functions instantiating the pizza application, the development of basic functions and subdialogues, digit recognition, tone input from telephone keypads, and alpha-digit recognition. Later on, grammars are created, and speech output, the use of TCL, and connection to a database are described. In Chapter 8, the CSLU Toolkit is used to develop a multimodal dialogue system. This includes the design of an animated character and a login dialogue as well as the consideration of emotions. The dialogue development using VoiceXML is shown in Chapter 9 and Chapter 10. On the basis of the tutorial, an introduction to VoiceXML is given and the reader learns how to integrate prompts, responses, verification, subdialogues, and tone/digit/alpha-digit recognition in a dialogue system. Moreover, the application of mixed initiative, the form interpretation algorithm (FIA), recognition grammars, variables, and Web server applications are dealt with. All the tutorials contain a detailed explanation of the respective topic, one or more examples for illustration, and, of course, multiple exercises based on these examples and explanations. The second part concludes with a detailed overview of XHTML+Voice and SALT, both of which are designed to enable multimodal access to Web-based services. Several examples of how to develop an XHTML+Voice-based application using the IBM Multimodal Toolkit and how to develop SALT-based systems with the help of the Microsoft .NET Speech SDK (Version 2.0 beta) are illustrated in Chapter 11.

In the third part, McTear presents advanced applications and technology. Chapter 12 describes advanced dialogue systems that involve more-complex tasks and provide sophisticated means to control the human-machine interaction; these are, for example, the DARPA Communicator, the TRAINS/TRIPS system, and the Conversational Architecture Project at Microsoft. Research topics in spoken-dialogue technology are discussed in Chapter 13. These topics include the information-state theory, error handling, adaptive dialogue systems, and the optimization of dialogue strategies. Chapter 14 shows future prospects of dialogue technology, especially multimodal systems.

McTear's book provides both a very recommendable entry point for novices or students who want to become familiar with spoken dialogue systems and a good reference manual for practitioners and researchers. The book is attractively presented and structured in a well-thought-out way. It is clearly written and easily understandable. The large number of examples given in the text introduce some variety to the topics. In addition to the standard references at the end of each chapter, Chapters 1–6 conclude

with an exercises section, in which the interested reader can delve into the topic. These exercises either contain questions about the content of the chapter or involve Web sites that demonstrate state-of-the-art technology. Although all the Web sites mentioned in these sections exist at the moment, it would have been good if the author or the publishing house had set up a central Web site for the book that would be updated constantly, taking into account changes to the referenced Web sites. Chapters 7–11 provide many tutorials, exercises, and examples that offer a big playing field for exploring the features and functionality of the respective toolkits and architectures. Instructions on how to obtain and install the necessary toolkits and software are given in the appendices. Some readers may miss a detailed list of abbreviations; however, most of the terms appear in the index.

So if your bank account were already capable of being articulate, it would probably say: “You possess more than \$59.95. Wouldn’t you like to buy *Spoken Dialogue Technology* by Michael F. McTear?”

Johannes Pittermann is a research assistant in the Dialogue Systems Group, Department of Information Technology, University of Ulm, Germany. His research focuses on adaptive dialogue management as well as speech signal preprocessing. His address is Department of Information Technology, Albert-Einstein-Allee 43, 89081 Ulm, Germany; e-mail: jpitterm@it.e-technik.uni-ulm.de.