Preface

In recent years spoken language research has been successful in establishing technology which can be used in various applications, and which has also brought forward novel research topics that advance our understanding of the human speech and communication processes in general. This book got started in order to collect these different trends together, and to provide an overview of the current state of the art, as well as of the challenging research topics that deal with spoken language interaction technologies.

The topics have been broadly divided into two main classes: research on the enabling technology on one hand and applications that exemplify the use of the technology on the other hand. Each chapter is an independent review of a specific topic, covering research problems and possible solutions, and also envisaging development of the field in the near future. The basic technology development covers areas such as automatic speech recognition and speech synthesis, spoken dialogue systems and dialogue modelling, expressive speech synthesis, emotions and affective computing, multimodal communication and animated agents, while the applications concern speech translation, spoken language usage in cars, space, and military applications, as well as applications for special user groups. Discussion of the general evaluation methodologies is also included in the book.

The authors are leading figures of their field. Their experience provides a strong basis for the discussion of various aspects of the specific research topic and, in addition to their own work, for the presentation of a broad view of the entire research area. The core topics are discussed from the research, engineering, and industrial perspectives, and together the chapters provide an integrated view of the research and development in speech and spoken dialogue technology, representing a comprehensive view of the whole area.

The structure of the contributions follows a general format that consists of presenting the current state of the art in the particular research area, including a short history and development of the research as the background, and then discussing and evaluating new trends and view points related to the future of the field. Concerning different applications, the articles address both technical and usability issues, e.g. they provide an analysis of the factors that affect the application's real-time functioning, and usability requirements concerning their specific interaction and design issues.

The book is intended for readers, both in industry and academia, who wish to get a summary of the current research and development in the speech technology and its related fields. The aim of the book is to highlight main research questions and possible solutions so as to clarify the existing technological possibilities for the design and development of spoken language interactive systems, and also to inspire experimentation of new techniques, methods, models, and applications in speech technology and interaction studies. It also serves as a reference book to the historical development of the field and its various subfields, and as a collection of future visions of the research directions.

The spoken interaction research field has made great advances during the past years, and is active, maybe even more than ever, with a wide range of research topics. Many challenges need to be addressed, and new topics will obviously come across during the investigations. As a part of the joint efforts in research and development, the book will, hopefully, help the development of technologies to go forward, so as to satisfy the requirements of the users, and also our needs to understand human communication better.

The book consists of 15 chapters which are divided into three sections: basic speech interaction technology, developments of the basic technology, and applications of spoken language technology. The chapters are briefly introduced below.

The first four chapters discuss the major research areas in speech and interaction technology: speech recognition, speech synthesis, and dialogue systems.

Sadaoki Furui surveys the major themes and advances of the speech recognition research in his chapter *History and Development of Speech Recognition*. The chapter provides a technological perspective of the five decades of speech research that has produced models and solutions for automatic speech recognition. The chapter also points out performance challenges under a broad range of operating conditions, and that a greater understanding of the human speech process is needed before speech recognition applications can approach human performance.

An overview of the speech synthesis research is given by **David Suendermann, Harald Höge** and **Alan Black** in their chapter *Challenges in Speech Synthesis*. In the first part, within a historical time scale that goes back about 1,000 years, various types of speaking machines are discussed, from the first mechanical synthesizers to the modern electronic speech synthesizers. The second part concerns speech corpora, standardisation, and the evaluation metrics, as well as the three most important techniques for the speech synthesis. The chapter finishes with the speech synthesis challenges, referring to evaluation competitions in speech synthesis that are believed to boost research and development of the field.

Kristiina Jokinen focuses on dialogue management modelling in her chapter *Spoken Language Dialogue Models*, and outlines four phases within the history of dialogue research, growing from one another as an evolutionary chain according to interests and development lines typical for each period. The chapter reviews the development from the point of view of a Thinking Machine, and also surveys different dialogue models that have influenced dialogue management systems and aimed at enabling natural interaction in speech and interaction technology.

Roberto Pieraccini discusses spoken dialogue technology from the industrial view point in his chapter *The Industry of Spoken Dialogue Systems and the Third Generation of Interactive Applications*. The starting point of the overview is in the change of perspective in spoken language technology, from the basic research of human-like natural understanding of spontaneous speech to technologically feasible solutions and development of dialogue systems. The chapter describes a general architecture of dialogue systems and also compares visual and voice applications. With a reference to spoken dialogue industry, the life cycle of speech applications is presented and the challenges of the third generation of spoken dialogue systems discussed.

The next five chapters move on to describe novel research areas which have grown within the main research and development work in recent years.

Julia Hirschberg presents an overview of the recent computational studies in spoken and written deception. Her chapter *Deceptive Speech: Clues from Spoken Language* reviews common approaches and potential features which have been used to study deceptive speech. Moreover, she reports some results on deception detection using language cues in general and spoken cues in particular.

Roger K. Moore discusses the cognitive perspective to speech communication in his chapter *Cognitive Approaches to Spoken Language Technology*. The chapter argues that, although spoken language technology has successfully migrated from the research laboratories into practical applications, there are shortfalls in the areas in which human beings excel, and that they would seem to result from the absence of cognitive level processes in contemporary systems. There is relatively little spoken language technology research that draws directly on models of human cognition or exploits the cognitive basis of human spoken language. The chapter attempts to redress the balance by offering some insights into where to draw insights and how this might be achieved.

Nick Campbell addresses the issue of human speech communication in his chapter *Expressive Speech Processing and Prosody Engineering.* He does not focus upon the linguistic aspects of speech, but rather on its structure and use in interactive discourse, and shows that prosody functions to signal much more than syntactic or semantic relationships. After considering prosodic information exchange from a theoretical standpoint, he discusses acoustic evidence for the ideas and finally suggests some technological applications that the broader view of spoken language interaction may give rise to.

Elisabeth André and **Catherine Pelachaud** concentrate on the challenges that arise when moving from speech-based human–computer dialogue to face-to-face communication with embodied conversational agents. Their chapter *Interacting with Embodied Conversational Agents* illustrates that researchers working on embodied conversational agents need to address aspects of social communication, such as emotions and personality, besides the objectives of robustness and efficiency that the work on speech-based dialogue is usually driven by. The chapter reviews ongoing research in the creation of embodied conversational agents and shows how these agents are endowed with human-like communicative capabilities:

the agents can talk and use different discourse strategies, display facial expressions, gaze pattern, and hand gestures in occurrence with their speech.

Jianhua Tao presents affective computing history and problem setting in his chapter *Multimodal Information Processing for Affective Computing*. The chapter studies some key technologies which have been developed in recent years, such as emotional speech processing, facial expression, body gesture, and movement, affective multimodal system, and affect understanding and generation. The chapter also introduces some related projects and discusses the key topics which comprise a large challenge in the current research.

The last five chapters of the book focus on applications, and it contains articles on different types of speech applications, including also a chapter on system evaluation.

Farzad Ehsani, Robert Frederking, Manny Rayner, and **Pierrette Bouillon** give a survey of speech-based translation in their chapter *Spoken Language Translation*. The chapter covers history and methods for speech translation, including the dream of a Universal Translator, and various approaches to build translation engines. The chapter discusses the specific requirements for speech translation engines, concerning especially the component technologies of the speech recognition and speech synthesis. Finally, some representative systems are introduced and their architecture and component technologies discussed in detail.

Fang Chen, Ing-Marie Jonsson, Jessica Villing, and **Staffan Larsson** in their, chapter *Application of speech technology in vehicles* summarise challenges of applying speech technology into vehicles, discuss the complexity of vehicle information systems, requirements for speech-based interaction with the driver, and discuss speech as an input/output modality. The chapter also presents dialogue-based conversational systems and multimodal systems as new next-level speech interfaces, and discusses the effects of emotion, mood, and the driver's personality on the application design.

Manny Rayner, Beth Ann Hockey, Jean-Michel Renders, Nikos Chatzichrisafis, and **Kim Farrel** describe Clarissa, a voice-enabled procedure browser which is apparently the first spoken dialogue system used in the International Space Station. The chapter *Spoken Dialogue Application in Space* focuses on the main problems and their solutions in the design and development of the Clarissa system. Detailed presentations are given on how to create voice-navigable versions of formal procedure documents, and how to build robust side-effect free dialogue management for handling undo's, corrections, and confirmations. The chapter also outlines grammar-based speech recognition using the Regulus toolkit and methods for accurate identification of cross-talk based on Support Vector Machines.

Jan Noyes and **Ellen Haas** describe various speech applications in a very demanding context: military domain. Their chapter *Military Applications* introduces the characteristics of the military domain by considering the users, technologies, and the environment in which the applications are used, pointing out harsh requirements for accuracy, robustness, and reliability in safety-critical conditions where applications are used in noisy and physically extreme environments by users who are subject to stress, time pressure, and workload. The chapter also presents some

typical speech-based applications dealing with Command-and-Control, teleoperation, information entry and information retrieval, repair and maintenance, training, and language translation.

Diamantino Freitas summarizes the use of speech technology in an alternative, or we may also say: augmentative way, to enable accessibility in communication technology for people with special needs. Problems as well as solutions are presented concerning specific situations of the visually disabled, the mobility impaired, the speech impaired, the hearing impaired, and the elderly. Problems found generally in public sites or sites providing transportation information require special attention in order to allow easy navigation and access to information. Also instructional games and eBooks are considered in examining what main benefits can be extracted from the use of speech technology in communication technology. It is concluded that solutions to improve communication difficulties for disabled persons may also bring advantages for non-disabled persons by providing redundancy and therefore a higher comfort in the use of communication systems.

Sebastian Möller summarizes the work on the evaluation of speech-based systems in his chapter *Assessment and Evaluation of Speech-based Interactive Systems*. After a brief history of the evaluation tasks, the chapter defines the concepts of performance and quality, which are often used to describe how well the service and the system fulfills the user's or the system designer's requirements. The chapter then moves on to a more detailed analysis of the assessment of individual system components: speech recognition, understanding, dialogue management, and speech synthesis, as well as the system as a whole. It also discusses a number of evaluation criteria that have been used to quantify the evaluation parameters, and which are partially standardized, as well as methods and frameworks for user evaluation. It is pointed out that new principles have to be worked out for capturing user experience and for evaluating multimodal systems, preferably in an automatic way.

Göteborg, Sweden — Fang Chen
Helsinki, Finland — Kristiina Jokinen

http://www.springer.com/978-0-387-73818-5

Speech Technology
Theory and Applications
Chen, F.; Jokinen, K. (Eds.)
2010, XXVII, 331 p. 188 illus., 23 illus. in color.,
Hardcover
ISBN: 978-0-387-73818-5