

Spoken Dialogue Technology

Springer-Verlag London Ltd.

Michael F. McTear

Spoken Dialogue Technology

Toward the Conversational User Interface

Foreword by T. V. Raman

 Springer

Michael F. McTear
School of Computing and Mathematics
University of Ulster
Newtownabbey BT37 0QB
Northern Ireland

British Library Cataloguing in Publication Data

McTear, Michael F., 1943–

Spoken dialogue technology : toward the conversational user interface

1. Natural language processing (Computer science) 2. User interfaces (Computer systems)

I. Title

006.3'5

Library of Congress Cataloging-in-Publication Data

McTear, Michael.

Spoken dialogue technology : toward the conversational user interface / Michael F. McTear.

p. cm.

Includes bibliographical references and index.

ISBN 978-1-85233-672-1 ISBN 978-0-85729-414-2 (eBook)

DOI 10.1007/978-0-85729-414-2

1. User interfaces (Computer systems) 2. Automatic speech recognition. I. Title.

QA76.9.U83M45 2004

005.4'37 – dc22

2004045619

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

Springeronline.com

© Springer-Verlag London 2004

Originally published by Springer-Verlag London Limited in 2004

The use of registered names, trademarks etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Typeset by SNP Best-set Typesetter Ltd., Hong Kong
34/3830-543210 Printed on acid-free paper SPIN 10886181

Foreword

The present coming of age of speech technologies coincides with the advent of mobile computing and the accompanying need for ubiquitous information access. This has generated enormous commercial interest around deploying speech interaction to IT-based services.

In his book, Michael gives an in-depth review of the nuts and bolts of constructing speech applications. Notice that creating good speech applications requires more than robust speech technologies; the creation of usable, efficient spoken interaction requires substantial thought toward crafting the man-machine conversation. This book does an excellent job in this regard by focusing on the creation of dialog-based applications.

As speech interaction enters the mainstream Web, these applications need to be developed, deployed and maintained against an efficient cost model. The speech industry is converging on a set of open XML-based standards developed under the auspices of the W3C in order to achieve this goal. VoiceXML, SSML and SRGF together form the underpinnings of the XML-based W3C Voice Web, and this book gives the reader a good foundation on building speech interaction using these interoperable standards. The closing chapters on multimodal interaction gives the reader the necessary background to start actively following and contributing to this exciting area.

T. V. Raman
Human Language Technologies
IBM Research

Preface

Writing this book has been a long process of discovery and learning that dates back to my first explorations in human–computer dialogue using natural language, documented in my book *The Articulate Computer* (Blackwell). Published in 1987, it described research in dialogue modeling where the primary input mode was the keyboard rather than speech. Since then things have come a long way. The fields of natural language processing and speech technology have come together, and the discipline of spoken dialogue technology has emerged. Moreover, there has been an increasing commercial interest in the technology, with a number of large, influential companies actively involved in spoken language technology and offering advanced toolkits and development platforms to researchers wishing to explore the exciting potential of conversing with a computer using spoken natural language.

Along the way I have met many people who have helped me and shaped my work. My first attempts with spoken dialogue interfaces involved acquiring and learning how to use the CSLU toolkit. This brought me into contact with Ron Cole, a pioneer in the development and promotion of freely available, open source software to support teachers and students wishing to learn about spoken language technologies. Through Ron I met several of his colleagues: Jacques de Villiers, Ed Kaiser and John-Paul Hosom at the Center for Spoken Language Understanding, who provided me with invaluable advice and support, and later, when Ron moved to the Center for Spoken Language Research (CSLR), Bryan Pellom and Wayne Ward, who introduced me to the CU Communicator and other toolkits and platforms being developed at CSLR.

More recently, as I started to learn about VoiceXML, I have found Jim Larson to be a tremendous source of inspiration and encouragement. As I moved on to look at more recent developments, such as XHTML + Voice, I found support and advice from a number of people in IBM's Pervasive Computing Group, most notably, T.V. Raman, David Jaramillo, Nick Metianu and Les Wilson. There are also a number of colleagues, who I meet regularly at conferences and other academic occasions and who have been a constant source of ideas. These include Harald Aust, Norman Fraser, Paul Heisterkamp, Alex Rudnicky, Paul McKeivitt, Ian O'Neill, David Toney and Colleen Crangle.

Developing the practical chapters of the book would not have been possible without the help of many colleagues and students. At the University of Ulster Stephen Downey spent many hours helping me to get the early versions of the CSLU toolkit installed and running properly and provided untiring support both to myself and to the students we taught together in lab sessions. Lesley-Ann Black has worked with me more recently on VoiceXML practicals and she also provided useful feedback on earlier drafts of the book. Countless students, too many to name, have been subjected to practicals involving the CSLU toolkit, VoiceXML, X+V, and SALT. I would like to single out Elizabeth Rooney and Angela Murphy, on whose Master's dissertations I have drawn extensively in chapters 6, 7 and 8, and also Conor McCluskey, Garth Kennedy and Aislinn McAleer, whose excellent undergraduate dissertations provided me with many useful ideas throughout. I would also like to acknowledge Gerrit Bloothoof and his students at the Utrecht Institute of Linguistics, who provided useful feedback and corrections on chapters 7 and 8 involving the CSLU toolkit, as well as students at the ELSNET summer school at Lille, 2004, who also worked through these chapters with me.

I was encouraged to write this book by Beverley Ford, Editorial Director of Springer-Verlag London. Jean Lovell-Butt provided me with editorial assistance during the writing of the book, and Louise Farkas, Senior Production Editor, Springer-Verlag New York, oversaw the development of the book from the raw typescript that I submitted to the final version of the book. To Beverley, Jean and Louise, and all of their colleagues who helped along the way, I offer my gratitude.

Finally, I must acknowledge my wife Sandra, who has supported and encouraged me throughout, particularly during the long hours away from family life that I have spent working on the book.

Newtownabbey, Northern Ireland

Michael F. McTear

Contents

Foreword by T. V. Raman	v
Preface	vii
Introduction	1
Who Should Read This Book?	2
How This Book Is Organised	2
Part I. Background to Spoken Dialogue Technology	5
1. Talking with Computers: Fact or Fiction?	7
A Little Fantasy	7
Further Reading	17
Exercises	18
2. Spoken Dialogue Applications: Research Directions and Commercial Deployment	19
Spoken Dialogue Applications	19
Research in Spoken Dialogue Technology	24
The Commercial Deployment of Spoken Dialogue Technology ...	37
Summary	41
Further Reading	42
Exercises	42
3. Understanding Dialogue	45
Dialogue: A Definition	45
Key Characteristics of Dialogue	46
Modelling Dialogue: Structure and Process	59
Summary	75
Further Reading	76
Exercises	77
4. Components of a Spoken Dialogue System – Speech Input and Output	79
Overview of a Spoken Dialogue System	79
Speech Recognition	81
Language Understanding	91

Language Generation	99
Text-to-Speech Synthesis	102
Summary	104
Further Reading	105
Exercises	105
5. Dialogue Management	107
Dialogue Initiative	108
Dialogue Control	111
Grounding	117
Accessing Information from an External Knowledge Source	121
Knowledge Sources for Dialogue Management	123
Issues for Developers	124
Summary	126
Further Reading	126
Exercise	126
Part II. Developing Spoken Dialogue Applications	127
6. Dialogue Engineering: The Dialogue Systems	
Development Lifecycle	129
Requirements Analysis	129
Requirements Specification	135
Design	139
Implementation	146
Testing	150
Evaluation	154
Summary	160
Further Reading	160
Exercises	161
7. Developing a Spoken Dialogue System Using the CSLU Toolkit ...	163
Tutorial 1: Basic Functions in RAD: The Pizza Application	164
Tutorial 2: Developing Basic Functions for a Student Information System	171
Tutorial 3: The Default Repair Subdialogue	174
Tutorial 4: Subdialogues	176
Tutorial 5: Digit Recognition	179
Tutorial 6: DTMF Input	181
Tutorial 7: Alpha-digit Recognition	184
Tutorial 8: Recognition Grammars	187
Tutorial 9: Speech Output in RAD	191
Tutorial 10: Using Tcl in RAD	192
Tutorial 11: Adding New Pronunciations and Creating a Dynamic Recogniser	194
Tutorial 12: Linking a RAD Application to a Database Using TclODBC	197
Summary	201

Further Reading	201
Ideas for Further Projects	202
8. Developing Multimodal Dialogue Systems Using the CSLU Toolkit	205
Tutorial 1: Setting Up the Animated Character	206
Tutorial 2: Login Dialogue	208
Tutorial 3: Exploring Emotions	211
Tutorial 4: Identifying Two-dimensional Shapes	217
Tutorial 5: Improving the Interface	224
Some Additional Features	227
Current Developments	229
Further Reading	229
Ideas for Further Projects	230
9. Developing a Directed Dialogue System Using VoiceXML	231
VoiceXML: A Language for Voice Enabling the World Wide Web	231
VoiceXML: The Language	233
Tutorials	233
Tutorial 1: Prompts, Responses and Verification	234
Tutorial 2: Repair in VoiceXML	241
Tutorial 3: Subdialogues	243
Tutorial 4: Digit Recognition	247
Tutorial 5: DTMF	249
Tutorial 6: Alphadigit Recognition	253
Tutorial 7: Modifying Speech Output	254
Tutorial 8: Navigation and Global Commands	257
Notes on ECMAScript	260
Further Reading	261
10. More Advanced VoiceXML	263
Tutorial 1: Mixed-initiative Dialogue	263
Tutorial 2: The Form Interpretation Algorithm (FIA)	266
Tutorial 3: Recognition Grammars	270
Tutorial 4: Variable and Grammar Scope	274
Tutorial 5: Dynamic VoiceXML and Web Server Applications	278
Tutorial 6: Dynamic Grammars	284
Further Reading	288
Ideas for Further Projects	288
11. Multimodal Web-based Dialogue: XHTML + Voice and SALT	291
XHTML + Voice	291
SALT	304
Summary	318
Further Reading	319
Ideas for Further Projects	320

Part III. Advanced Applications	321
12. Advanced Dialogue Systems: Some Cases Studies	323
DARPA Communicator	323
TRAINS and TRIPS	332
Conversational Architectures Project	341
Summary	347
Further Reading	348
13. Research Topics in Spoken Dialogue Technology	349
Information State Theory	349
Error Handling in Spoken Dialogue Systems	353
Adaptive Dialogue Systems	362
Learning Optimal Dialogue Strategies	371
Summary	375
Further Reading	376
14. Future Directions: The Way Ahead	379
Advanced Multimodal Systems	379
Future Directions	385
Concluding Remarks	388
Appendices	389
Appendix 1. Grammar Formats	391
The XML and ABNF Grammar Formats	391
Appendix 2. The CSLU Toolkit and Related Software for Chapters 7 and 8	395
CSLU Toolkit	395
TclODBC	395
Setting Up a Data Source (ODBC)	395
Appendix 3. Software for Chapters 9 and 10	397
Downloading the IBM WebSphere VoiceServer SDK	397
Downloading and Installing Apache Tomcat	399
Java Server Pages (JSP)	399
Appendix 4. Software for Chapter 11. IBM Multimodal Tools and Microsoft.NET Speech SDK	400
IBM Multimodal Tools	400
Microsoft .NET Speech SDK	400
Appendix 5. Spoken Dialogue Technology: Projects and Links	401
Appendix 6. Spoken Dialogue Technology: Toolkits and Resources ...	406
Toolkits and Resources for Spoken Dialogue Technology	406
Web Sites for Hosted VoiceXML Applications	407
References	409
Index	425