

T E X T , S P E E C H  
A N D L A N G U A G E  
T E C H N O L O G Y

---

# PRACTICAL SPOKEN DIALOG SYSTEMS

Edited by Deborah Dahl

---

Academic Publishers

TN912.3  
p895

# Practical Spoken Dialog Systems

Edited by

Deborah Dahl

*Conversational Technologies, Plymouth Meeting, U.S.A.*



E200501479

**KLUWER ACADEMIC PUBLISHERS**

DORDRECHT / BOSTON / LONDON

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 1-4020-2674-9 (HB)  
ISBN 1-4020-2676-5 (e-book)

---

Published by Kluwer Academic Publishers,  
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Sold and distributed in North, Central and South America  
by Kluwer Academic Publishers,  
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed  
by Kluwer Academic Publishers,  
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

*Printed on acid-free paper*

All Rights Reserved

© 2004 Kluwer Academic Publishers

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed in the Netherlands

# Practical Spoken Dialog Systems

# Text, Speech and Language Technology

---

VOLUME 26

---

## *Series Editors*

Nancy Ide, *Vassar College, New York*

Jean Véronis, *Université de Provence and CNRS, France*

## *Editorial Board*

Harald Baayen, *Max Planck Institute for Psycholinguistics, The Netherlands*

Kenneth W. Church, *AT & T Bell Labs, New Jersey, USA*

Judith Klavans, *Columbia University, New York, USA*

David T. Barnard, *University of Regina, Canada*

Dan Tufis, *Romanian Academy of Sciences, Romania*

Joaquim Llisterri, *Universitat Autònoma de Barcelona, Spain*

Stig Johansson, *University of Oslo, Norway*

Joseph Mariani, *LIMSI-CNRS, France*

*The titles published in this series are listed on [www.wkap.nl/prod/s/TLTB](http://www.wkap.nl/prod/s/TLTB).*

*For Rich, Sarah, and Peter*

## **PREFACE**

Deployed spoken dialog systems, in which humans and computers engage in a conversation using speech, are becoming increasingly commonplace. This book includes papers on a variety of topics organized around the theme of what it really takes to build a successful spoken dialog system. Working spoken dialog systems need to take into account both principles of human-computer communication as well as the considerable fund of practical knowledge that is developing in the industry as more and more of these applications are developed and used. The goal of this book is to provide a resource for managers, developers, and students who want to gain insight into both the communication principles of human-computer dialog systems as well as the details of what it takes to design, develop, and deploy a working spoken dialog system. In providing this information, this book fills the gap between introductory material and the research literature on human-computer interaction. It is hoped that the reader will take from this book a deeper and fuller understanding of the practical considerations required to turn basic technologies such as speech recognition, dialog design principles, VoiceXML, and design tools into running systems. In addition to understanding techniques for developing running systems based on current technologies, it's also valuable to look beyond current technology and to understand what kinds of new capabilities are coming along in the near future, and this book also contains papers that will spark the readers' imaginations in that direction.

Specific topics covered in the papers in this book include the business aspects of deciding how a speech application will address business needs, several advanced design topics, an in-depth look at how a specific application works in practice, and examples of near-term research work.

Deborah A. Dahl

## CONTRIBUTORS

Samuel Bayer, *The Mitre Corporation*

Deborah A. Dahl, *Conversational Technologies*

Morena Danieli, *Loquendo Vocal Technologies*

Kurt Godden, *General Motors R&D*

Li Gong, *SAP Labs*

Robert Keiller, *VoxSurf*

James A. Larson, *Intel Corporation*

Judith A. Markowitz, *J. Markowitz, Consultants*

Clifford Nass, *Stanford University*

Candace L. Sidner, *Mitsubishi Research Laboratories*

Suzanne Liebowitz Taylor, *Unisys Corporation*



# TABLE OF CONTENTS

|                    |      |
|--------------------|------|
| PREFACE .....      | xi   |
| CONTRIBUTORS ..... | xiii |
| INTRODUCTION.....  | 1    |

## PART 1: INITIAL STAGES

|  |    |
|--|----|
| SUZANNE LIEBOWITZ TAYLOR / Beyond Technology ..... | 7  |
| 1. Introduction.....                               | 7  |
| 2. Speech Engagements.....                         | 9  |
| 3. Discovery .....                                 | 13 |
| 4. Summary .....                                   | 22 |

## PART 2: DESIGN

|   |    |
|---|----|
| CLIFFORD NASS AND LI GONG / Ten Principles for<br>Designing Human-Computer Dialog Systems ..... | 25 |
| 1. Introduction.....  | 25 |
| 2. Processing Speech.....   | 27 |
| 3. Producing Speech.....  | 33 |
| 4. Spoken Dialog Interaction.....   | 36 |
| 5. Final Thoughts .....   | 38 |

|  |    |
|--|----|
| JAMES A. LARSON / Voice User Interface Design for<br>Novice and Experienced Users..... | 41 |
| 1. Introduction.....   | 41 |
| 2. Developing Voice User Interfaces .....  | 47 |
| 3. Guidelines For Voice User Interfaces for Novice Users.....                          | 47 |
| 4. Guidelines for Voice User Interfaces for Average Users.....                         | 52 |
| 5. Guidelines for User Interfaces for Experienced Users.....                           | 54 |
| 6. Advanced Techniques .....   | 60 |
| 7. Iterative Testing .....   | 62 |
| 8. Summary .....   | 63 |

|   |     |
|---|-----|
| MORENA DANIELI / Designing Error Recovery Dialogs ..... | 65  |
| 1. Introduction .....                                   | 65  |
| 2. Human-machine Miscommunication .....                 | 67  |
| 3. Kinds of Miscommunication .....                      | 70  |
| 4. The Design of Error Recovery Subdialogs .....        | 74  |
| 5. How Users Signal Misunderstanding .....              | 81  |
| 6. Conclusions .....                                    | 81  |
| DEBORAH A. DAHL / Visualization Tools for Designing     |     |
| Spoken Dialogs .....                                    | 85  |
| 1. Problem .....  | 85  |
| 2. Visualizing Dialogs .....                            | 86  |
| 3. The Role of Tools in Development .....               | 87  |
| 4. Example Dialog .....                                 | 88  |
| 5. General Issues and Selection criteria .....          | 102 |
| 6. Conclusion .....                                     | 103 |
| 7. Acknowledgements .....                               | 104 |
| KURT GODDEN / How to Wreck a Nice Speech Grammar .....  | 105 |
| 1. Introduction .....                                   | 105 |
| 2. Limitations of Speech Grammars .....                 | 106 |
| 3. Who is the User? .....                               | 108 |
| 4. Grammar Design .....                                 | 114 |
| 5. Error Prompts .....                                  | 116 |
| 6. Complexity .....                                     | 117 |
| 7. Ambiguity .....                                      | 118 |
| 8. Miscellaneous .....                                  | 120 |
| 9. Conclusions .....                                    | 121 |
| JUDITH A. MARKOWITZ / Designing for Speaker             |     |
| Authentication .....                                    | 123 |
| 1. Introduction .....                                   | 123 |
| 2. Speaker Authentication Dialogs .....                 | 126 |
| 3. Advanced Techniques .....                            | 135 |
| 4. Handling Errors and Problems .....                   | 137 |
| 5. Final Words .....                                    | 139 |

## PART 3: DEPLOYMENT

|   |     |
|---|-----|
| ROBERT KEILLER / Using VoiceXML 2.0 in the VxOne Unified Messaging Application..... | 143 |
| 1. Introduction.....  | 144 |
| 2. Overview of VxOne.....   | 144 |
| 3. Generation of Dynamic VoiceXML pages.....  | 148 |
| 4. Speed and Latency.....   | 157 |
| 5. Prompt Generation.....   | 159 |
| 6. Conclusions.....   | 161 |

## PART 4: NEW IDEAS

|   |     |
|---|-----|
| SAMUEL BAYER / Building a Standards and Research Community with the Galaxy Communicator Software Infrastructure ..... | 167 |
| 1. Introduction.....  | 167 |
| 2. Enabling an Engineering Community.....   | 168 |
| 3. The GALAXY Communicator software infrastructure.....   | 172 |
| 4. Design.....  | 176 |
| 5. Implementation and Logistics.....  | 189 |
| 6. Standards, Commercialization and the GCSI.....   | 191 |
| 7. Conclusion.....  | 194 |

|  |     |
|--|-----|
| CANDACE L. SIDNER / Building Spoken-Language Collaborative Agents..... | 197 |
| 1. Introduction.....   | 197 |
| 2. Collaborative Interface Agents.....                                 | 199 |
| 3. Four Spoken Language Collaborative Agents.....                      | 205 |
| 4. Design Issues for Collaborative Agents.....                         | 219 |
| 5. Related Work.....   | 224 |
| 6. A Recap of Lessons Learned.....                                     | 224 |

|            |     |
|------------|-----|
| INDEX..... | 227 |
|------------|-----|

# INTRODUCTION

The number of deployed spoken dialog systems has exploded in the last few years. In hundreds of applications running all over the world, people can now speak to computers and have them perform useful tasks — make travel reservations, provide news or weather information, read email — based only on the users' speech. These spoken dialog systems combine the technologies of automatic speech recognition, speech synthesis, natural language processing and dialog control to create effective spoken human-computer dialog systems.

The recent explosion of systems is due to a combination of several factors — greatly improved speech recognition technology, the vision and energy of the entrepreneurs who've taken the risk of bringing this speech technology to market, and most recently, the development of standards such as VoiceXML (Voice eXtensible Markup Language) [1] that dramatically lower the barrier to application development.

One of VoiceXML's most appealing features is that it takes very little time to grasp the basic ideas and start generating simple speech applications. There are a number of good ways to get this basic knowledge, including books such as [2, 3] and the VoiceXML specification itself. But these exploratory applications are a long way from deployed, 24/7 systems running in the field.

This book is aimed at addressing the question of what it means to go beyond these first applications, not only for applications built in VoiceXML, but for speech applications in general. Certainly one way of going beyond simple applications is to build more complex applications, with more functionality and more complex back-end integration, such as integration with the web and with back-end databases. There are very good books that can help the developer move forward in this direction, for example, [4, 5]. There are also very good existing books that can help designers create more sophisticated and user-friendly designs by understanding the principles of human factors of spoken dialog such as [3, 6, 7].

This book tackles other dimensions of spoken dialog applications. What this book does is start filling in the gap between the kinds of simple applications that can be built in a few hours, and effective, practical, dialog applications that solve real problems for customers. There are many issues around this topic, and this book is only a start. The papers in this book discuss the issues of the importance of understanding the customer's needs and expectations regarding speech applications (Taylor) advanced design issues (Larson, Nass, Danieli, Godden), incorporation of new technologies like speaker verification (Markowitz), the role of design tools (Dahl),

practical issues in deployment of systems using VoiceXML in particular (Keiller) and ideas for future directions in dialog technology (Bayer, Sidner). We focus here on what can be accomplished with both today's technology, and with technology that will be available within the next few years. Because it is focused on current and near-term technology, the book doesn't attempt to cover the rich research literature on spoken dialog systems, although many of the individual papers include pointers into this literature which the interested reader can follow to explore the research issues in more depth.

The book is organized roughly around the application development life cycle, from initial customer discussions through to the details of a fully deployed system, and then concludes with two papers looking to the future. Because VoiceXML is a popular and well-known dialog design framework, many of the papers use VoiceXML to provide very concrete examples of the principles they discuss. However, the principles are applicable to spoken human-computer dialogs in general, regardless of the implementation technology.

Part 1 discusses the initial stages of a project, before design has even begun, with Suzanne Taylor's paper on working with customers in the initial stages of a project. As in any project, technical soundness can't guarantee success if the vendor and customer have different expectations about what the application is and why it's being done. Taylor lays out in detail the numerous considerations of business drivers, environment, customer readiness, and requirements, that have to be gotten right at the outset in order for a project to be successful.

Once the initial selection of the project is completed, design begins. Part 2 covers a wide range of design issues, including the specialized topics of designing speech recognition grammars and speaker authentication systems. Part 2 begins with Clifford Nass and Li Gong's paper on ten important principles based on the evolution of human communication capabilities that have to be taken into account to build effective dialog systems. They make a compelling and insightful claim: Because speech evolved over millions of years for human-to-human communication, it includes many properties that need to be accommodated in understanding human to computer communication.

Most spoken dialog systems have to accommodate both novice and experienced users. A system designed only for novice users will bore experienced users, but novices will be confused by a system designed only for experienced users. James Larson's paper provides guidelines for designing systems for novice, average, and experienced users, as well as techniques for accommodating a spectrum of user capabilities in a single system.

Errors are inevitable in spoken dialog systems. Misrecognitions by automatic speech recognition systems are an obvious source of errors, but as Morena Danieli points out in her paper, many errors originate from systems which can't handle user's natural behavior in error recovery situations. Danieli provides ten best practices for adding effective error recovery capabilities to dialog systems.

When the time comes to put the principles from the papers by Nass and Gong, Larson, and Danieli into practice, developers will look for tools that can assist them in this complex task. My own paper discusses a number of approaches to visualizing spoken dialogs that have been incorporated into existing dialog development tools, or which might be incorporated into future tools. This information will be very useful in the process of selecting design tools.

Designing spoken dialog systems includes designing not only the dialog itself, but also the grammars that support automatic speech recognition. This is a complex task that is extremely important to a successful speech system. Kurt Godden's paper provides eight very useful guidelines for designing good speech grammars.

The last paper in Part 2 is Judith Markowitz's paper on designing for speaker authentication. Many applications need to restrict their access to specific users. Speaker authentication technologies make it possible to base this restriction on the speaker's voice. In addition to providing guidelines for the use of speaker authentication technologies in dialog systems, Markowitz's paper also provides a good introduction to these technologies that will be useful for those who are less familiar with this topic.

Part 3 covers deployment of spoken dialog systems, with Robert Keiller's case study of a specific system implemented by VoxSurf using VoiceXML. He describes the VoxSurf application as well as many important issues that arise in deployment of a full-scale spoken dialog system. These include runtime issues such as speed and latency as well as issues that arise in the use of other technologies, such as output pre-processing for speech synthesis systems. Keiller concludes with some suggestions for improvements to future versions of VoiceXML.

In Part 4 we move beyond today's currently deployed systems and take a look at some future directions in dialog systems. While today's deployed systems typically use either a proprietary architecture or VoiceXML, there are also other innovative open architectures resulting from research efforts. Samuel Bayer's paper discusses the Galaxy Communicator system. The Communicator program provides a framework for integrating research contributions from different sites so that each site can focus on its own area of expertise.

As impressive as today's systems are compared to the systems which were available only a few years ago, there are still significant areas for improvement. Today's systems don't have a high level understanding of the task that the user is attempting to accomplish, so they aren't capable of assisting users in the achievement of their goals. The final paper in the book, Candace Sidner's paper on collaborative interface agents, provides an intriguing picture of sophisticated dialogs with a collaborative agent that can do much of the detailed planning which has to be done by the user in current systems.

I hope you find the papers in this book as useful, thought provoking, and insightful as I do.

Deborah A. Dahl

## REFERENCES

- [1] W3C, "Voice Extensible Markup Language (VoiceXML 2.0)  
<http://www.w3.org/TR/voicexml20/>", <http://www.w3.org/TR/voicexml20/>, 2002.
- [2] M. Miller, *VoiceXML: 10 Projects to Voice-Enable Your Web Site*. New York: Wiley Publishing, Inc., 2002.
- [3] J. A. Larson, *VoiceXML: Introduction to developing speech applications*. Upper Saddle River New Jersey: Prentice Hall, 2002.
- [4] E. A. Andersson, S. Breitenbach, T. Burd, N. Chidambaram, P. Houle, D. Newsome, X. Tang, and X. Zhu, *Early Adopter VoiceXML*. Birmingham, UK: Wrox Press, 2001.
- [5] C. Sharma and J. Kunins, *VoiceXML*. New York: John Wiley and Sons, Inc., 2002.
- [6] B. Balentine and D. Morgan, *How to build a speech recognition application*. San Ramon, California: Enterprise Integration Group, 1999.
- [7] D. Gardner-Bonneau, *Human Factors and Voice Interactive Systems*. Boston: Kluwer Academic Publishers, 1999.

## PART 1

### INITIAL STAGES



