

SpringerBriefs in Speech Technology

Series Editor:

Amy Neustein

For other titles published in this series, go to

<http://www.springer.com/series/10043>

Editor's Note

The authors of this series have been hand-selected. They comprise some of the most outstanding scientists – drawn from academia and private industry – whose research is marked by its novelty, applicability, and practicality in providing broad based speech solutions. The SpringerBriefs in Speech Technology series provides the latest findings in speech technology gleaned from comprehensive literature reviews and *empirical investigations* that are performed in both laboratory and *real life* settings. Some of the topics covered in this series include the presentation of real life commercial deployment of spoken dialog systems, contemporary methods of speech parameterization, developments in information security for automated speech, forensic speaker recognition, use of sophisticated speech analytics in call centers, and an exploration of new methods of soft computing for improving human-computer interaction. Those in academia, the private sector, the self service industry, law enforcement, and government intelligence, are among the principal audience for this series, which is designed to serve as an important and essential reference guide for speech developers, system designers, speech engineers, linguists and others. In particular, a major audience of readers will consist of researchers and technical experts in the automated call center industry where speech processing is a key component to the functioning of customer care contact centers.

Amy Neustein, Ph.D., serves as Editor-in-Chief of the International Journal of Speech Technology (Springer). She edited the recently published book "Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics" (Springer 2010), and serves as quest columnist on speech processing for Womensenews. Dr. Neustein is Founder and CEO of Linguistic Technology Systems, a NJ-based think tank for intelligent design of advanced natural language based emotion-detection software to improve human response in monitoring recorded conversations of terror suspects and helpline calls. Dr. Neustein's work appears in the peer review literature and in industry and mass media publications. Her academic books, which cover a range of political, social and legal topics, have been cited in the Chronicles of Higher Education, and have won her a pro Humanitate Literary Award. She serves on the visiting faculty of the National Judicial College and as a plenary speaker at conferences in artificial intelligence and computing. Dr. Neustein is a member of MIR (machine intelligence research) Labs, which does advanced work in computer technology to assist underdeveloped countries in improving their ability to cope with famine, disease/illness, and political and social affliction. She is a founding member of the New York City Speech Processing Consortium, a newly formed group of NY-based companies, publishing houses, and researchers dedicated to advancing speech technology research and development.

David Suendermann

Advances in Commercial Deployment of Spoken Dialog Systems

 Springer

David Suendermann
SpeechCycle, Inc.
26 Broadway 11th Floor
New York, NY 10004
USA
david@speechcycle.com

ISSN 2191-737X e-ISSN 2191-7388
ISBN 978-1-4419-9609-1 e-ISBN 978-1-4419-9610-7
DOI 10.1007/978-1-4419-9610-7
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011930670

© Springer Science+Business Media, LLC 2011

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Spoken dialog systems have been the object of intensive research interest over the past two decades, and hundreds of scientific articles as well as a handful of text books such as [25, 52, 74, 79, 80, 83] have seen the light of day. What most of these publications lack, however, is a link to the “real world”, i.e., to conditions, issues, and environmental characteristics of deployed systems that process millions of calls every week resulting in millions of dollars of cost savings. Instead of learning about:

- Voice user interface design.
- Psychological foundations of human-machine interaction.
- The deep academic¹ side of spoken dialog system research.
- Toy examples.
- Simulated users.

the present book investigates:

- Large deployed systems with thousands of activities whose calls often exceed 20 min of duration.
- Technological advances in deployed dialog systems (such as reinforcement learning, massive use of statistical language models and classifiers, self-adaptation, etc.).
- To which extent academic approaches (such as statistical spoken language understanding or dialog management) are applicable to deployed systems – if at all.

¹This book draws a line between core research on spoken dialog systems as performed in academic institutions and in large industrial research labs on the one hand and commercially deployed spoken dialog systems on the other hand. As a convention, the former will be referred to as *academic*, the latter as *deployed* systems.

To Whom It May Concern

There are three main statements touched upon above:

1. Huge commercial significance of deployed spoken dialog systems.
2. Lack of scientific publications on deployed spoken dialog systems.
3. Overwhelming difference between academic and deployed systems.

These arguments, further backed up in Chap. 1, indicate a strong need for a comprehensive overview about the state of the art in deployed spoken dialog systems. Accordingly, major topics covered by the present book are as follows:

- After a brief introduction to the general architecture of a spoken dialog system, Chap. 1 offers some insight into important parameters of deployed systems (such as traffic, costs) before *comparing the worlds of academic and deployed spoken dialog systems* in various dimensions.
- *Architectural paradigms* for all the components of deployed spoken dialog systems are discussed in Chap. 2. This chapter will also deal with the many limitations deployed systems face (with respect to e.g. functionality, openness of input/output language, performance) imposed by hardware requirements, legal constraints, and the performance and robustness of current speech recognition and understanding technology.
- The key to success or failure of deployed spoken dialog systems is their performance. Performance being a diffuse term when it comes to the (continuous) *evaluation of dialog systems*, Chap. 3 will be dedicated to why, what, and when to measure performance of deployed systems.
- After setting the stage for a continuous performance evaluation, the logical consequence is trying to increase system performance on an ongoing basis. This attempt is often realized as a continuous cycle involving multiple *techniques for adapting and optimizing* all the components of deployed spoken dialog systems as discussed in Chap. 4. Adaptation and optimization are essential to deployed applications because of two main reasons:
 1. Every application can only be suboptimal when deployed for the first time due to the absence of live data during the initial design phase. Hence, application tuning is crucial to make sure deployed spoken dialog systems achieve maximum performance.
 2. Caller behavior, call reasons, caller characteristics, and business objectives are subject to change over time. External events that can be of irregular (such as network outages, promotions, political events), seasonal (college football season, winter recess), or slowly progressing nature (slow migration from analog to digital television, expansion of the Smartphone market) may have considerable effects on what type of calls an application must be able to handle.

Due to the book's focus on paradigms, processes, and techniques applied to deployed spoken dialog systems, it will be of primary interest to speech scientists,

voice user interface designers, application engineers, and other technical staff of the automated call center industry, probably the largest group of professionals in the speech and language processing industry. Since Chap. 1 as well as several other parts of the book aim at bridging the gap between academic and deployed spoken dialog systems, the community of academic researchers in the field is in focus as well.

New York City
February 2011

David Suendermann

Acknowledgements

The name of the series which the present book is a volume of, SpringerBriefs, makes use of two words that have a meaning in the German language: Springer (knight) and Brief (letter). Indeed, I was fighting hard like a knight to get this letter done in less than four months of sleepless nights. In this effort, several remarkable people stood by me: Dr. Amy Neustein, Series Editor of the SpringerBriefs in Speech Technology, whose strong editing capabilities I learned to greatly appreciate in a recent similar project, kindly invited me to author the present monograph. Essential guidance and support in the course of this knight ride came also from the editorial team at Springer – Alex Greene and Andrew Leigh. On the final spurt, Dr. Roberto Pieraccini as well as Dr. Renko Geffarth contributed invaluable reviews of the entire volume adding the finishing touches to the manuscript.

Contents

1	Deployed vs. Academic Spoken Dialog Systems	1
1.1	At-a-Glance	1
1.2	Census, Internet, and a Lot of Numbers	2
1.3	The Two Worlds.....	7
2	Paradigms for Deployed Spoken Dialog Systems	9
2.1	A Few Remarks on History.....	9
2.2	Components of Spoken Dialog Systems	11
2.3	Speech Recognition and Understanding	12
2.3.1	Rule-Based Grammars.....	12
2.3.2	Statistical Language Models and Classifiers	13
2.3.3	Robustness	14
2.4	Dialog Management	25
2.5	Language and Speech Generation	27
2.6	Voice Browsing	30
2.7	Deployed Spoken Dialog Systems are Real-Time Systems.....	33
3	Measuring Performance of Spoken Dialog Systems	39
3.1	Observable vs. Hidden	39
3.2	Speech Performance Analysis Metrics.....	42
3.3	Objective vs. Subjective	46
3.4	Evaluation Infrastructure	48
4	Deployed Spoken Dialog Systems' Alpha and Omega: Adaptation and Optimization	49
4.1	Speech Recognition and Understanding	50
4.2	Dialog Management	55
4.2.1	Escalator.....	55

4.2.2 Engager.....	56
4.2.3 Contender	59
References	63