Advances in Pattern Recognition

Advances in Pattern Recognition is a series of books which brings together current developments in all areas of this multi-disciplinary topic. It covers both theoretical and applied aspects of pattern recognition, and provides texts for students and senior researchers.

Springer also publishes a related journal, **Pattern Analysis and Applications**. For more details see: http://springlink.com

The book series and journal are both edited by Professor Sameer Singh of Loughborough University, UK.

Also in this series:

Principles of Visual Information Retrieval
Michael S. Lew (Ed.)
1-85233-381-2

Statistical and Neural Classifiers: An Integrated Approach to Design
Šarūnas Raudys
1-85233-297-2

Advanced Algorithmic Approaches to Medical Image Segmentation
Jasjit Suri, Kamaledin Setarehdan and Sameer Singh (Eds)
1-85233-389-8

NETLAB: Algorithms for Pattern Recognition
Ian T. Nabney
1-85233-440-1

Object Recognition: Fundamentals and Case Studies
M. Bennamoun and G.J. Mamic
1-85233-398-7

Computer Vision Beyond the Visible Spectrum
Bir Bhanu and Ioannis Pavlidis (Eds)
1-85233-604-8

Hexagonal Image Processing: A Practical Approach
Lee Middleton and Jayanthi Sivaswamy
1-85233-914-4

Support Vector Machines for Pattern Classification
Shigeo Abe
1-85233-929-2

Digital Document Processing
Bidyut B. Chaudhuri (Ed.)
978-1-84628-501-1

Recent Advances in Pattern Recognition
Sameer Singh and Maneesha Singh (Eds)
978-1-84628-944-6

Human Ear Recognition by Computer
Bir Bhanu and Hui Chen
978-1-84800-128-2

Zheng-Hua Tan and Børge Lindberg

Automatic Speech Recognition on Mobile Devices and over Communication Networks

Springer

Zheng-Hua Tan, BSc, MSc, PhD
Børge Lindberg, MSc
Department of Electronic Systems, Aalborg University, Aalborg, Denmark

Series editor
Professor Sameer Singh, PhD
Research School of Informatics, Loughborough University, Loughborough, UK

Advances in Pattern Recognition Series ISSN 1617-7916
ISBN-13: 978-1-84800-142-8 e-ISBN-13: 978-1-84800-143-5
DOI 10.1007/978-1-84800-143-5

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2008920253

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

Springer Science+Business Media
springer.com

Preface

The remarkable advances in computing and networking have sparked an enormous interest in deploying *Automatic Speech Recognition on Mobile Devices and Over Communication Networks*, and the trend is accelerating. This yields an abundance of practical systems, operational algorithms and scientific publications. There is, however, no integrated book available that portrays the whole picture of this area. Our primary impetus for editing this book is to fill this gap by providing a comprehensive and unified introduction to the field.

The prevalence of mobile devices, coupled with the proliferation of wireless networks, creates new opportunities for speech recognition technology. Mobile devices are small in size and are used while on the move, both of which make speech-enabled user interfaces attractive in comparison with other interaction modes like keypad and stylus. The opportunities come along with challenges as well. For instance, it is not an easy task to port state-of-the-art speech recognition systems onto computationally limited devices such as mobile phones, PDAs and automobiles where they are highly desirable. Fortunately, the barriers are being removed because of increasingly powerful embedded platforms and pervasive network connections. Still, however, the accompanying research and engineering issues are many: computational constraints and power limitations on the devices, speech coding and transmission deteriorations over the networks, diverse operating systems and hardware configurations, to name just a few. To address these issues requires a wide scope of knowledge and experience.

This book brings together leading researchers and practitioners from academia and industry to provide an in-depth review of methods and standards, share working knowledge, and present state-of-the-art systems and applications. We cover network speech recognition, distributed speech recognition and embedded speech recognition, which are expected to co-exist in the coming years.

Organization and Features

The book begins with an overview chapter and is then divided into four parts: network speech recognition, distributed speech recognition, embedded speech recognition, and systems and applications.

Chapter 1 gives a comprehensive overview of network, distributed and embedded speech recognition and discusses the pros and cons of the presented approaches. This chapter sets the scene for the entire book.

Part I, **Network Speech Recognition**, focuses on remote speech recognition that uses conventional speech coders for the transmission of speech from a client device to a recognition server where feature extraction and recognition decoding take place. This part consists of three chapters.

Chapter 2 first describes the commonly used speech coding standards for mobile and IP networks, and then investigates the effect of speech codecs on speech and speaker recognition performance, with or without packet loss. Chapter 3 addresses issues related to speech recognition over mobile networks, and presents solutions to the performance degradation caused by speech coding algorithms, transmission errors and environmental noise. Chapter 4 reviews robustness techniques against packet loss in the context of voice over IP-based network speech recognition, and introduces a CELP-type speech coder optimized for speech recognition over IP networks.

Part II, **Distributed Speech Recognition**, makes a thorough presentation of speech recognition that adopts the client-server architecture by placing feature extraction in the client and recognition decoding in the server. It begins with a review of distributed speech recognition standards. The subsequent four chapters cover the major blocks of distributed speech recognition.

Chapter 5 provides a comprehensive overview of the industry standards for distributed speech recognition developed in ETSI, 3GPP and IETF in addition to a summary of substantial performance testing and comparisons to AMR coded speech. Chapter 6 presents techniques for feature extraction and back-end speech reconstruction from the MFCC features on the basis of voicing and fundamental frequency information either transmitted from the client device or predicted from the received features. Chapter 7 describes a series of schemes for quantizing the MFCC features, including scalar quantization, vector quantization and block quantization, where the optimization objective is to maximize recognition accuracy. Chapter 8 presents a survey of error recovery methods for transmitting the quantized features over error-prone channels, including both forward error control coding that adds redundancy to the feature stream and interleaving that creates spread in it. Client-side error recovery cannot completely prevent the occurrence of residual bit errors or packet loss. Chapter 9 therefore concentrates on sever-side error concealment to reduce the detrimental effect induced by transmission errors.

Part III, **Embedded Speech Recognition**, addresses the main problems in realizing a speech recognition system fully on a mobile device. The problems are approached from both algorithm and arithmetic sides through three dedicated chapters.

Chapter 10 presents an overview of algorithm implementations and optimizations aimed at a speech recognition system with a low computational complexity and thus suitable for deployment on embedded platforms. To complement this, Chap. 11 primarily targets a low memory footprint and emphasizes on techniques for compressing HMMs by removing redundancies from HMMs through parameter tying and state- or density-clustering and by quantizing HMMs. Chapter 12 reviews problems concerning the fixed-point arithmetic implementation of speech recognition algorithms and presents fixed-point methods that give the same recognition accuracy as that of floating-point algorithms.

Part IV, **Systems and Applications**, introduces practical work and knowledge. It starts with the introduction to architecture considerations in a network environment. The succeeding three chapters present speech recognition systems and applications tailored for mobile phones, PDAs and automobiles, respectively. The last chapter presents energy-aware speech recognition for mobile devices.

Chapter 13 examines software architectures for mobile speech applications from an industrial viewpoint with a thorough comparison between embedded and distributed speech engines and a highlight on supporting multimodal user interaction. Chapter 14 presents applications of speech recognition for mobile phones and puts the focuses on multilinguality, noise robustness, and footprint and complexity reduction. Chapter 15 presents a two-way free-form speech-to-speech translation system that includes a large vocabulary continuous speech recognizer, a translation module and a multi-language speech synthesis system and is completely hosted on a PDA. Chapter 16 describes the development of speech technology components for various automotive applications and reviews issues and challenges related to automotive platforms. With a concern that battery technology significantly lags behind semiconductor technology, Chap. 17 investigates the system-level energy consumption from both computation and communication of distributed speech recognition on a wireless device and presents a set of optimization algorithms that can increase the battery lifetime by an order of magnitude.

A comprehensive **index** is provided at the end of this book. Index words are highlighted in the text by using italic font.

While chapters are complemented to each other and are presented in a unified manner with a clear flow from chapter to chapter, each chapter is written to be self-contained and can be read and understood independently. As such, certain redundancy is kept in the book. The book contains chapters of a tutorial nature as well as chapters on research advances and practical applications.

Target Audiences

The book is primarily intended for students, engineers and scientists working in speech processing and recognition. This book can also be a reference for practitioners and researchers involved in user interface and application design for mobile devices, speech communication over networks, Internet and wireless communications, and data compression.

Supplementary Materials

For more information about software, databases, literature and related links, please refer to the book's Web site, http://asr.es.aau.dk.

Acknowledgements

We warmly thank the authors, our friends and colleagues, for their outstanding contributions and for their hard work in making a timely, high-quality publication of this book a reality. We are grateful to Wayne Wheeler, Catherine Brett, Beverley Ford and Helen Callaghan at Springer for their great support and assistance.

Zheng-Hua Tan
Børge Lindberg

Aalborg, Denmark

Contents

Preface .... **v**

Contributors .... **xix**

1. Network, Distributed and Embedded Speech Recognition: An Overview **1**

Zheng-Hua Tan and Imre Varga

1.1 Introduction 1
1.2 ASR and Its Deployment in Devices and Networks 3
1.2.1 Automatic Speech Recognition 3
1.2.2 Resources and Constraints of Mobile Devices 5
1.2.3 Resources and Constraints of Communication Networks 7
1.2.4 Architectural Solutions for ASR in Devices and Networks 8
1.3 Network Speech Recognition 9
1.4 Distributed Speech Recognition 11
1.4.1 Feature Extraction 11
1.4.2 Source Coding 12
1.4.3 Channel Coding and Packetisation 13
1.4.4 Error Concealment 14
1.4.5 DSR Standards 14
1.4.6 A Configurable DSR System 15
1.5 Embedded Speech Recognition 15
1.5.1 ESR Scenario 16
1.5.2 Applications and Platforms 16
1.5.3 Fixed-Point Arithmetic 17
1.5.4 Optimisation 18
1.5.5 Robustness 19
1.6 Discussion 20
References 21

Part I Network Speech Recognition

2. Speech Coding and Packet Loss Effects on Speech and Speaker Recognition 27
Laurent Besacier
2.1 Introduction 27
2.2 Sources of Degradation in Network Speech Recognition 28
2.2.1 Speech and Audio Coding Standards 28
2.2.2 Packet Loss 30
2.3 Effects on the Automatic Speech Recognition Task 32
2.3.1 Experimental Setup 32
2.3.2 Degradation Due to Simulated Packet Loss 32
2.3.3 Degradation with Real Transmissions 33
2.3.4 Degradation Due to Speech and Audio Codecs 34
2.4 Effect for the Automatic Speaker Verification Task 35
2.4.1 Speaker Verification Experiments Over Compressed Speech and Packet Loss 36
2.4.2 Speaker Verification Experiments Over GSM Compressed Speech 37
2.5 Conclusion 38
Acknowledgments 38
References 39

3. Speech Recognition Over Mobile Networks 41
Hong Kook Kim and Richard C. Rose
3.1 Introduction 41
3.2 Techniques for Improving ASR Performance Over Mobile Networks 43
3.3 Bitstream-Based Approach 46
3.4 Feature Transform 50
3.4.1 Mel-Scaled LPCC 51
3.4.2 LPC-Based MFCC (LP-MFCC) 52
3.4.3 Pseudo-Cepstrum (PCEP) and Its Mel-Scaled Variant (MPCEP) 53
3.5 Enhancement of ASR Performance Over Mobile Networks 53
3.5.1 Compensation for the Effect of Mobile Systems 53
3.5.2 Compensation for Speech Coding Distortion in LSP Domain 54
3.5.3 Compensation for Channel Errors 56
3.6 Conclusion 57
References 58

4. Speech Recognition Over IP Networks 63
Hong Kook Kim
4.1 Introduction 63
4.2 Speech Recognition and IP Networks 65
4.2.1 Relationship Between ASR Performance and Speech Quality 65
4.2.2 Impact of Speech Coding Distortion 66
4.2.3 Impact of Network Channel Distortion 67

4.3 Robustness Against Packet Loss 69
4.3.1 Rate Control 69
4.3.2 Forward Error Correction 70
4.3.3 Interleaving 70
4.3.4 Error Concealment and ASRDecoder-Based Concealment 71
4.4 Speech Coder for Speech Recognition Over IP Networks 71
4.4.1 MFCC-Based Speech Coder 72
4.4.2 Efficient Vector Quantization of MFCCs 74
4.4.3 Speech Quality Comparison 78
4.4.4 ASR Performance Comparison 79
4.5 Conclusion 82
References 82

Part II Distributed Speech Recognition

5. Distributed Speech Recognition Standards 87
David Pearce
5.1 Introduction 87
5.2 Overview of the Set of DSR Standards 89
5.3 Scope of the Standards 90
5.3.1 Electro-Acoustics 91
5.3.2 Speech Detection or External Control Signal 92
5.3.3 Pre-Processing 92
5.3.4 Parameterisation 92
5.3.5 Compression and Error Protection 93
5.3.6 Formatting 93
5.3.7 Error Detection and Mitigation 93
5.3.8 Decompression 93
5.3.9 Server Side Post Processing 93
5.3.10 Feature Derivatives 93
5.4 DSR Basic Front-End ES 201 108 94
5.4.1 Feature Extraction 94
5.4.2 Compression 94
5.4.3 Error Detection and Mitigation 95
5.5 DSR Advanced Front-End ES 202 050 96
5.5.1 Feature Extraction 96
5.5.2 VAD 96
5.5.3 Compression 96
5.6 Recognition Performance of the DSR Front-Ends 97
5.6.1 Aurora Speech Databases and ETSI performance Testing 97
5.6.2 Aurora 3: Multilingual SpeechDat-Car Digits — Small Vocabulary Evaluation 97
5.7 3GPP Evaluations and Comparisons to AMR Coded Speech 99
5.8 ETSI DSR Extended Front-End Standards ES 202 211 and ES 202 212 102
5.9 Transport Protocols: The IETF RTP Payload Formats for DSR 104
5.10 Conclusion 105
Acknowledgements 105
References 105

6. Speech Feature Extraction and Reconstruction 107
Ben Milner
6.1 Introduction 107
6.2 Feature Extraction 109
6.2.1 Basic Terminal-Side Feature Extraction 109
6.2.2 Advanced Terminal-Side Feature Extraction 115
6.2.3 Quantisation and Packetisation 116
6.2.4 Server-Side Processing 117
6.3 Speech Reconstruction 117
6.3.1 Analysis of Received Speech Information 118
6.3.2 Speech Reconstruction 119
6.4 Prediction of Voicing and Fundamental Frequency 123
6.4.1 Fundamental Frequency Prediction from MFCC Vectors 123
6.4.2 Voicing Prediction from MFCC Vectors 126
6.4.3 Speech Reconstruction from Predicted Fundamental Frequency and Voicing 128
6.5 Conclusion 129
References 129

7. Quantization of Speech Features: Source Coding 131
Stephen So and Kuldip K. Paliwal
7.1 Introduction 131
7.2 Quantization Schemes 132
7.2.1 Brief Introduction to Quantization Theory 132
7.2.2 Distortion Measures for Quantization in Speech Processing 134
7.2.3 Scalar Quantization 135
7.2.4 Block Quantization 137
7.2.5 Vector Quantization 137
7.2.6 GMM-Based Block Quantization 138
7.3 Quantization of ASR Feature Vectors 141
7.3.1 Introduction and Literature Review 141
7.3.2 Statistical Properties of MFCCs 142
7.3.3 Use of Cepstral Liftering for MFCC Variance Normalization 148
7.3.4 Relationship Between the Distortion Measure and Recognition Performance 150
7.3.5 Improving Noise Robustness: Perceptual Weighting of Filterbank Energies 152
7.4 Experimental Results 153
7.4.1 ETSI Aurora-2 Distributed Speech Recognition Task 153
7.4.2 Experimental Setup 154
7.4.3 Non-Uniform Scalar Quantization Using HRO Bit Allocation 154
7.4.4 Unconstrained Vector Quantization 155
7.4.5 GMM-Based Block Quantization 156
7.4.6 Multi-frame GMM-Based Block Quantization 156
7.4.7 Perceptually-Weighted Vector Quantization of Logarithmic Filterbank Energies 157
7.5 Conclusion 158
References 159

8. Error Recovery: Channel Coding and Packetization 163
Bengt J. Borgström, Alexis Bernard, and Abeer Alwan
8.1 Distributed Speech Recognition Systems 163
8.2 Characterization and Modeling of Communication Channels 164
8.2.1 Signal Degradation Over Wireless Communication Channels....... 164
8.2.2 Signal Degradation Over IP Networks 165
8.2.3 Modeling Bursty Communication Channels.................................. 165
8.3 Media-Specific FEC.. 167
8.4 Media-Independent FEC 168
8.4.1 Combining FEC with Error Concealment Methods........................ 169
8.4.2 Linear Block Codes 169
8.4.3 Cyclic Codes.. 174
8.4.4 Convolutional Codes 174
8.5 Unequal Error Protections 176
8.6 Frame Interleaving 177
8.6.1 Optimal Spread Block Interleavers.. 178
8.6.2 Convolutional Interleavers 179
8.6.3 Decorrelated Block Interleavers 180
8.7 Examples of Modern Error Recovery Standards.................................... 181
8.7.1 ETSI DSR Standard (ETSI 2000).. 181
8.7.2 ETSI GSM/EFR Standard (ETSI 1998)... 182
8.8 Summary 183
Acknowledgements.. 184
References.. 184

9. Error Concealment 187
Reinhold Haeb-Umbach and Valentin Ion
9.1 Introduction... 187
9.2 Speech Recognition in the Presence of Corrupted Features................... 190
9.2.1 Modified Observation Probability 190
9.2.2 Gaussian Approximation 193
9.3 Feature Posterior Estimation in a DSR Framework 194
9.3.1 ETSI DSR Standards 195
9.3.2 Source Coder Redundancy 195
9.3.3 Channel Models... 196
9.3.4 Estimation of Feature Posterior 199
9.3.5 Related Work... 201
9.4 Performance Evaluations 202
9.4.1 Experimental Setup 202
9.4.2 Results on GSM Data Channel.. 203
9.4.3 Results on Packet Erasure Channel 206
9.5 Conclusion 207
Acknowledgments.. 208
References.. 208

Part III Embedded Speech Recognition

10. Algorithm Optimizations: Low Computational Complexity 213
Miroslav Novak
10.1 Introduction 213
10.2 Common Limitations of Embedded Platforms 214
10.2.1 Memory Limitations 214
10.2.2 CPU Limitations 215
10.3 Overview of an ASR System 215
10.4 Front End 216
10.5 Observation Model 217
10.5.1 Model Organization 217
10.5.2 Efficient Computation Strategies 218
10.6 Search 221
10.6.1 Viterbi Search Implementation 222
10.6.2 Search Graph Construction 226
10.6.3 Fast Match 228
10.6.4 Alternative Decoding Schemes 228
10.7 Conclusion 229
Acknowledgments 229
References 230

11. Algorithm Optimizations: Low Memory Footprint 233
Marcel Vasilache
11.1 Introduction 233
11.2 Notations and Problem Statement 234
11.3 Model Complexity Control 237
11.3.1 Akaike's Information Criterion 238
11.3.2 Bayesian Information Criterion 238
11.3.3 Second Order Approximation 239
11.3.4 Other Measures 239
11.4 Parameter Tying 239
11.4.1 Model Level 240
11.4.2 State Level 241
11.4.3 Density Level 241
11.4.4 Subspaces 241
11.4.5 Clustering 242
11.5 Parameter Representations 243
11.5.1 Floating Point Representation 243
11.5.2 Fixed Point Representation 244
11.5.3 Quantization 244
11.6 Quantized Parameters HMMs 245
11.6.1 Scalar Quantization 245
11.6.2 Vector Quantization 247
11.7 Subspace Distribution Clustering HMM 247
11.7.1 Subspace Partitioning 248
11.7.2 Density Clustering 249

11.8 Computational Complexity Implications 249
11.9 Practicalities and Conclusion 250
References 251

12. Fixed-Point Arithmetic 255
Enrico Bocchieri
12.1 Introduction 255
12.2 Fixed-Point Arithmetic 257
12.2.1 Programming with Fixed-Point Numbers 257
12.2.2 Fixed-Point Representation and Quantization 259
12.3 LVCSR MAP Recognizer 259
12.3.1 HMM State Likelihoods 261
12.3.2 State Duration Model 262
12.3.3 Language Model 263
12.3.4 Viterbi Decoder 263
12.3.5 Acoustic Front-End 264
12.4 Fixed-Point Implementation of the Recognizer 264
12.4.1 Log-Likelihoods 265
12.4.2 Viterbi Frame-Synchronous Search 266
12.4.3 Gaussian Parameters 267
12.4.4 MFCC Front-End 268
12.5 Experiments 269
12.5.1 Real-Time on the Device 272
12.6 Conclusion 274
Acknowledgements 274
References 274

Part IV Systems and Applications

13. Software Architectures for Networked Mobile Speech Applications 279
James C. Ferrans and Jonathan Engelsma
13.1 Introduction 279
13.1.1 Embedded and Distributed Speech Engines 279
13.1.2 The Voice Web 280
13.1.3 Multimodal User Interfaces 283
13.1.4 Distributed Speech Recognition 284
13.1.5 Multimodal Architectures 285
13.1.6 Simultaneous and Sequential Multimodality 287
13.1.7 Mode Composition 288
13.2 Classes of Multimodal Architectures 288
13.2.1 Fully Embedded or "Fat Client" (a) 289
13.2.2 Distributed Processing Engines (b) 289
13.2.3 Thin Client (d) 291
13.2.4 Remote Visual Interface (e) 291
13.2.5 "Pudgy" Client (c) 292
13.2.6 Discussion 292
13.3 The "Plus V" Distributed Multimodal Architecture 293
13.4 Other Distributed Multimodal Architectures 295

13.4.1 Video Interactive Services with VoiceXML 295
13.4.2 Multimodal for Set-Top Boxes... 295
13.4.3 Bare Minimum Mobile Voice Search.. 296
13.4.4 A Transcription-Based Architecture... 297
13.5 Towards a Commercial Ecosystem... 297
13.6 Conclusion 298
References.. 298

14. Speech Recognition in Mobile Phones.. 301
Imre Varga and Imre Kiss
14.1 Introduction... 301
14.2 Applications of Speech Recognition for Mobile Phones 302
14.3 Multilinguality and Language Support 305
14.3.1 Multilingual Speaker Independent Name Dialing 305
14.3.2 Multilinguality in Other ASR Applications............................... 308
14.3.3 Language Resources 308
14.4 Noise Robustness 309
14.4.1 Robust HMM Models... 309
14.4.2 Feature Extraction 309
14.4.3 Noise Reduction 310
14.5 Footprint and Complexity Reduction.. 314
14.5.1 Footprint Reduction of Acoustic Models.................................. 314
14.5.2 Footprint Reduction of Language Models................................ 315
14.5.3 Footprint Reduction of Pronunciation Lexicon 317
14.5.4 Reduction of Computational Complexity in Embedded ASR Systems 317
14.5.5 Low Memory, Fast Decoding... 319
14.6 Platforms and an Example Application.. 319
14.6.1 Example Application: Large Vocabulary Isolated Word Dictation 320
14.7 Conclusion and Outlook.. 323
References.. 323

15. Handheld Speech to Speech Translation System..................................... 327
Yuqing Gao, Bowen Zhou, Weizhong Zhu and Wei Zhang
15.1 Introduction... 327
15.2 System Overview 328
15.2.1 System architecture 328
15.2.2 Hardware and OS Specifications.. 330
15.2.3 Interface.. 330
15.3 System Components and Optimization... 332
15.3.1 LVCSR on Handheld Devices.. 332
15.3.2 Natural Language Understanding and Generation Based Translation.. 334
15.3.3 Weighted Finite State Transducer Based Translation 337
15.3.4 Embedded Speech Synthesis 340

15.4 Experiments and Discussions 341
15.4.1 Speech Recognition Experiments 341
15.4.2 Translation Experiments 343
15.5 Conclusion 344
References 345

16. Automotive Speech Recognition 347
Harald Höge, Sascha Hohenner, Bernhard Kämmerer, Niels Kunstmann, Stefanie Schachtl, Martin Schönle, and Panji Setiawan
16.1 Introduction 347
16.2 Siemens Speech Processing — From Research to Products 348
16.2.1 Development for Performance and Quality 348
16.2.2 High-Performance Recognizer 349
16.2.3 Ultra-Compact Text-to-Speech Synthesizer 350
16.2.4 Natural Voice Dialog 351
16.2.5 Speaker Characterization and Recognition 351
16.3 Example Automotive Voice Applications: Infotainment, Navigation, Manuals, and Internet 351
16.3.1 Radio Station Selection 352
16.3.2 MP3 Title Selection 352
16.3.3 Navigation Destination Entry 353
16.3.4 Manuals and Help Systems 354
16.3.5 Access to Structured Web Content 355
16.3.6 Access to Web Services 356
16.4 Automotive Platform Issues and Challenges 357
16.4.1 Hardware Constraints 358
16.4.2 Software Constraints 359
16.4.3 User Constraints 360
16.4.4 Acoustic Channel 360
16.5 Noise Robust Recognition Technology 360
16.5.1 ASR Front-End 362
16.5.2 Minimum Mean Square Weighting Rules 363
16.5.3 Recursive Least Squares Weighting Rules 364
16.5.4 Implementation of RLS Weighting Rules 365
16.5.5 Recognition Results 366
16.6 Methodology for Evaluation of Automotive Recognizers Quality Measurement Using SNR Curves 367
16.6.1 Common Evaluation Procedures 368
16.6.2 Proposed SNR-Approach 368
16.6.3 Data Recording 368
16.6.4 Evaluation 369
16.6.5 Best Practice 371
16.7 Conclusion 372
References 372

17. Energy Aware Speech Recognition for Mobile Devices 375
Brian Delaney
17.1 Introduction 375
17.1.1 Battery Technology 375
17.1.2 Energy Aware Design Principles 376
17.1.3 Related Work 377
17.2 Case Study of Distributed Speech Recognition Using the HP Labs Smartbadge System 379
17.2.1 Signal Processing Front-End 379
17.2.2 Energy Consumption of DSR with IEEE 802.11 Wireless Networks 384
17.2.3 Energy Consumption of DSR Using Bluetooth Networks 389
17.2.4 Comparison of 802.11 and Bluetooth in DSR 391
17.3 Conclusion 395
References 395

Index 397

Contributors

Abeer Alwan
University of California, Los Angeles,
Department of Electrical Engineering,
Los Angeles, CA, USA
alwan@ee.ucla.edu

Alexis Bernard
University of California, Los Angeles,
Department of Electrical Engineering,
Los Angeles, CA, USA
Bernard@UCLAlumni.net

Laurent Besacier
University J. Fourier,
LIG Laboratory,
Grenoble, France
Laurent.Besacier@imag.fr

Enrico Bocchieri
AT&T Labs Research, Florham Park,
New Jersey, USA
enrico@research.att.com

Bengt J. Borgström
University of California, Los Angeles,
Department of Electrical Engineering,
Los Angeles, CA, USA
jonas@ee.ucla.edu

Brian Delaney
Massachusetts Institute of Technology,
Lincoln Laboratory, Information Systems
Technology Group,
Lexington, MA, USA
bdelaney@ll.mit.edu

Jonathan Engelsma
Motorola Labs, USA
jonathan.engelsma@motorola.com

James C. Ferrans
Motorola Labs,
USA
james.ferrans@motorola.com

Yuqing Gao
IBM T. J. Watson Research Center,
USA
yuqing@us.ibm.com

Reinhold Haeb-Umbach
University of Paderborn,
Department of Communications
Engineering,
33095 Pad erborn, Germany
haeb@nt.uni-paderborn.de

Harald Höge
Siemens AG, Corporate Technology,
81739 München, Germany
harald.hoege@siemens.com

Sascha Hohenner
Siemens AG, Corporate Technology,
81739 München, Germany
sascha.hohenner@siemens.com

Valentin Ion
University of Paderborn,
Department of Communications
Engineering,
33095 Paderborn, Germany
ion@nt.uni-paderborn.de

Bernhard Kämmerer
Siemens AG, Corporate Technology,
81739 München, Germany
bernhard.kaemmerer@siemens.com

Hong Kook Kim
Gwangju Institute of Science
and Technology, Department
of Information and Communications,
Gwangju, Korea
hongkook@gist.ac.kr

Imre Kiss
Nokia, Finland
imre.kiss@nokia.com

Niels Kunstmann
Siemens AG, Corporate Technology,
81739 München, Germany
niels.kunstmann@siemens.com

Børge Lindberg
Aalborg University,
Department of Electronic Systems,
9220 Aalborg, Denmark
bli@es.aau.dk

Ben Milner
University of East Anglia,
School of Computing Sciences,
Norwich, NR4 7TJ, United Kingdom
b.milner@uea.ac.uk

Miroslav Novak
IBM T.J Watson Research Center,
Speech and Language Technologies,
USA
miroslav@us.ibm.com

Kuldip K. Paliwal
Griffith University,
Griffith School of Engineering,
Signal Processing Laboratory,
QLD 4222 Australia
k.paliwal@griffith.edu.au

David Pearce
Motorola Labs,
Applications Research Centre,
Basingstoke, UK
david.pearce@motorola.com

Richard C. Rose
McGill University, Department
of Electrical and Computer Engineering,
Montreal, Quebec, Canada
rose@ece.mcgill.ca

Stefanie Schachtl
Siemens AG, Corporate Technology,
81739 München, Germany
stefanie.schachtl@siemens.com

Martin Schönle
Siemens AG, Corporate Technology,
81739 München, Germany
martin.schoenle@siemens.com

Panji Setiawan
Universität der Bundeswehr München,
München, Germany
panji.setiawan@yahoo.de

Stephen So
Griffith University,
Griffith School of Engineering,
Signal Processing Laboratory,
QLD 4222 Australia
s.so@griffith.edu.au

Zheng-Hua Tan
Aalborg University,
Department of Electronic Systems,
9220 Aalborg, Denmark
zt@es.aau.dk

Imre Varga
Siemens AG,
Corporate Technology, Germany
imre.varga@nsn.com

Marcel Vasilache
Nokia, 33100 Tampere, Finland
marcel.vasilache@nokia.com

Wei Zhang
IBM T. J. Watson Research Center,
USA
zhangwei@us.ibm.com

Bowen Zhou
IBM T. J. Watson Research Center,
USA
zhou@us.ibm.com

Weizhong Zhu
IBM T. J. Watson Research Center,
USA
zhuwe@us.ibm.com