

## Table of Contents

### Foreword

Keith M. Ponting .....	VII
1. Insight vs. performance .....	VIII
2. Dangers .....	IX
3. Hot topics .....	IX
4. Towards the future .....	X
4.1 Integrating knowledge sources .....	XI
4.2 A unified theory? .....	XI
References .....	XII

### Speech Pattern Processing

Roger K. Moore .....	1
1. The State-of-the-Art in Speech .....	1
2. Speech Patterning .....	2
3. Speech Pattern Processing .....	3
4. Whither a Unified Theory? .....	5
4.1 Towards a Theory .....	5
4.2 Practical Issues .....	5
5. What We Know .....	6
6. Some Things We Don't Know .....	7
7. The Way Forward .....	7
References .....	8

### Psycho-acoustics and Speech Perception

Louis C.W. Pols .....	10
1. Introduction .....	10
2. Psycho-acoustics .....	11
3. Speech Perception .....	13
3.1 Vowel Reduction and Schwa .....	13
3.2 Spectro-temporal Dynamics of Formant Transitions .....	14
3.3 Consonant Reduction .....	14
4. Discussion .....	15
References .....	16

### **Acoustic Modelling for Large Vocabulary Continuous Speech Recognition**

Steve Young .....	18
1. Introduction .....	18
2. Overview of LVCSR Architecture .....	18
3. Front End Processing .....	21
4. Basic Phone Modelling .....	22
4.1 HMM Phone Models .....	22
4.2 HMM Parameter Estimation .....	24
4.3 Context-Dependent Phone Models .....	26
5. Adaptation for LVCSR .....	30
5.1 Maximum Likelihood Linear Regression .....	31
5.2 Estimating the MLLR Transforms .....	31
6. Progress in LVCSR .....	33
7. Discriminative Training for LVCSR .....	34
8. Conclusions .....	37
References .....	38

### **Tree-based Dependence Models for Speech Recognition**

Mari Ostendorf, Ashvin Kannan and Orith Ronen .....	40
1. Introduction .....	40
2. Hidden Tree Framework .....	41
3. Hidden Dependence Trees .....	43
3.1 The Mathematical Framework .....	43
3.2 Application to Speech .....	44
3.3 Topology Design and Parameter Estimation .....	44
3.4 Experiments .....	46
4. Multiscale Tree Processes .....	47
4.1 The Mathematical Framework .....	47
4.2 Application to Speech .....	48
4.3 Topology Design and Parameter Estimation .....	49
4.4 Experiments .....	50
5. Discussion .....	51
References .....	52

### **Connectionist and Hybrid Models for Automatic Speech Recognition**

Jean-Paul Haton .....	54
1. Introduction .....	54
2. A Brief Overview of Neural Networks .....	55
2.1 Basic Principles .....	55
2.2 Main Models for ASR .....	56
3. Signal Processing and Feature Extraction using ANNs .....	57
4. Neural Networks as Static Pattern Classifiers .....	58
4.1 Speech Pattern Classification with Perceptrons .....	58

4.2 FeatureMaps .....	58
5. Dynamic Aspects .....	59
5.1 Position of the Problem .....	59
5.2 Time Delays .....	59
5.3 Dynamic Classifiers .....	59
5.4 Recurrent NNs .....	60
6. Hybrid Models .....	61
6.1 Position of the Problem .....	61
6.2 Proposed Solutions .....	61
7. Conclusion .....	63
References .....	63

### **Computational Models for Auditory Speech Processing**

Li Deng .....	67
I. Introduction .....	67
2. A nonlinear computational model for basilar membrane wave motions .....	67
3. Frequency-domain and time-domain computational solutions to the BM model .....	68
4. Interval analysis of auditory model's outputs for temporal information extraction .....	70
5. IPIH representation of clean and noisy speech sounds .....	71
6. Speech recognition experiments .....	73
7. Summary and discussions .....	75
References .....	76

### **Speaker Adaptation of CDHMMs Using Bayesian Learning**

Claudio Vair and Luciano Fissore .....	78
1. Introduction .....	78
2. Bayesian Estimation of CDHMMs .....	78
2.1 Prior Density Definition .....	79
2.2 Forgetting Mechanism .....	79
2.3 Prior Parameter Estimation and MAP Solution .....	80
3. Acoustic Normalization .....	81
4. Tasks, Corpus and System .....	81
5. Speaker Adaptation Experiments .....	82
6. Conclusions .....	83
References .....	83

### **Discriminative Improvement of the Representation Space for Continuous Speech Recognition**

Ángel de la Torre, Antonio M. Peinado, Antonio J. Rubio, José C. Segura ... ..	84
1. Introduction .....	84
2. Discriminative Feature Extraction .....	84
3. SGDFE Algorithm for CSR .....	85

4. Experimental Results .....	86
5. Conclusions .....	89
References .....	89

### **Dealing with Loss of Synchronism in Multi-Band Continuous Speech Recognition Systems**

Christophe Cerisara .....	90
1. Introduction .....	90
2. Forcing Synchronism Between the Bands .....	91
2.1 First Approach .....	91
2.2 Experiments .....	92
3. Modeling Loss of Synchronism .....	92
3.1 Theoretical Approach .....	92
3.2 Experimental Approach .....	93
4. Conclusion .....	94
References .....	95

### **K-Nearest Neighbours Estimator in a HMM-Based Recognition System**

Fabrice Lefèvre, Claude Montacié and Marie-José Caraty .....	96
1. Introduction .....	96
2. K-NN Assessment .....	96
3. K-NN estimator in HMM .....	97
3.1 Adaptation Principle .....	98
3.2 HMM Estimation Improvement .....	98
4. Evaluations .....	99
4.1 Recognition rates .....	99
4.2 SNALC Evaluation .....	100
5. Perspectives .....	101
References .....	101

### **Robust Speech Recognition**

Sadaaki Furui .....	102
1. Mismatches between Training and Testing .....	102
1.1 Speech Variation .....	102
1.2 Inter-Speaker Variation .....	104
2. Reducing Mismatches to Improve Speech Recognition .....	104
2.1 Principles of Adaptive Speech Recognition .....	104
2.2 Three Principal Adaptation Methods for Reducing Mismatches .....	106
2.3 Important Practical Issues .....	107
2.4 N-Best-Based Unsupervised Adaptation .....	108
3. Conclusion .....	109
References .....	109

**Channel Adaptation**

Keith M. Ponting .....	112
I. Introduction .....	112
1.1 Matched condition training .....	112
1.2 Robust features .....	112
1.3 Model adaptation .....	113
1.4 Channel adaptation .....	113
1.5 Speech enhancement .....	113
2. Models of distortion .....	113
2.1 Minimum mean square error .....	114
2.2 Additive noise estimation .....	114
3. Methods for channel adaptation .....	115
3.1 Global transformations .....	115
3.2 Class-specific corrections .....	116
3.3 Empirical methods based on stereo data .....	117
3.4 Model-based compensation .....	118
4. Conclusion .....	119
References .....	120

**Speaker Characterization, Speaker Adaptation and Voice Conversion**

Sadaaki Furui .....	122
1. Introduction .....	122
2. Speaker-Characterization .....	122
3. Speaker Recognition .....	123
4. Speaker-Adaptation Techniques for Speech Recognition .....	124
4.1 Classification of Speaker-Adaptation/Normalization Methods .....	124
4.2 Speaker Cluster Selection Methods .....	124
4.3 Interpolated Re-Estimation Algorithm .....	125
4.4 Spectral Mapping Algorithm .....	125
5. Individuality Problems in Speech Synthesis and Coding .....	128
6. Conclusion .....	129
References .....	130

**Speaker Recognition**

Sadaaki Furui .....	132
1. Principles of Speaker Recognition .....	132
2. Text-Independent Speaker Recognition Methods .....	133
2.1 Long-Term-Statistics-Based Methods .....	133
2.2 VQ-Based Methods .....	135
2.3 Ergodic-HMM-Based Methods .....	135
2.4 Speech-Recognition-Based Methods .....	136
3. Text-prompted Speaker Recognition .....	137
4. Normalization and Adaptation Techniques .....	137
4.1 Parameter-Domain Normalization .....	138

4.2 Likelihood Normalization .....	138
4.3 HMM Adaptation for Noisy Conditions .....	139
4.4 Updating Models and A Priori Threshold for Speaker Verification ... ..	139
5. Open Questions and Concluding Remarks .....	140
References .....	140

### **Application of Acoustic Discriminative Trainmug in an Ergodic HMM for Speaker Identification**

Leandro Rodríguez Liñares and Carmen García Mateo .....	143
1. Introduction .....	143
2. Experimental Conditions .....	144
3. System Architecture .....	145
3.1 Acoustic Segmentation .....	145
3.2 The PTh-HMM Model .....	145
4. Experimental Results .....	145
5. Conclusions .....	147
References .....	148

### **Comparison of Several Compensation Techniques for Robust Speaker Verification**

Laura Docfo-Fernández and Carmen García-Mateo .....	149
1. Introduction .....	149
2. The HMM recognition system .....	151
3. Mismatch Compensation Techniques .....	151
3.1 CMS .....	151
3.2 5M1 .....	152
3.3 SM2 .....	152
4. Experiments and Results .....	152
5. Discussion and Conclusion .....	156
References .....	156

### **Segmental Acoustic Modeling for Speech Recognition**

Man Ostendorf .....	157
I. Introduction .....	157
2. Segmental and Hidden Markov Models .....	158
2.1 General Modeling Framework .....	159
2.2 Models of Feature Dynamics .....	161
3. Recognition and Training .....	165
3.1 Recognition Algorithms .....	165
3.2 Parameter Estitnation Algoritlitns .....	166
4. Segmental Features .....	168
5. Sununary .....	169
References .....	170

### **Trajectory Representations and Acoustic Descriptions for a Segment-Modelling Approach to Automatic Speech Recognition**

Wendy J. Holmes .....	173
1. Introduction .....	173
2. Modelling Trajectories in Speech .....	174
3. Representing an Unobserved Trajectory with Segmental HMMs .....	175
3.1 Calculating segment probabilities .....	175
3.2 Recognition experiment .....	176
4. HMM Recognition with Formant Features .....	177
5. Modelling trajectories of cepstrum and formant features .....	178
6. Conclusions .....	178
References .....	179

### **Suprasegmental Modelling**

E. Nöth, A. Batliner, A. Kießling, R. Kompe and H. Niemann .....	181
I. Introduction .....	181
2. The Verbmobil System .....	183
3. Computation of Prosodic Information .....	183
3.1 Extraction of Prosodic Features .....	185
3.2 Prosodic Classes .....	185
3.3 New Boundary Labels: The Syntactic-prosodic M-labels .....	186
3.4 Classification of Prosodic Events .....	187
3.5 Improving the Classification Results with Stochastic Language Models ..	187
3.6 Prosodic scoring of WHGs .....	189
4. The Use of Prosodic Information .....	190
4.1 Prosody and Syntax – Interaction with the TUG-Grammar .....	190
4.2 Prosody and the Other Linguistic Modules .....	194
5. Concluding Remarks .....	196
References .....	196

### **Computational Models for Speech Production**

Li Deng .....	199
1. Introduction .....	199
2. Speech production models in science/technology literatures .....	200
3. Derivation of discrete-time version of statistical task-dynamic model .....	202
4. Algorithms for learning task-dynamic model parameters and for likelihood computation .....	204
4.1 Model with deterministic, time-invariant parameters .....	205
4.2 Model with random, time-invariant parameters .....	207
4.3 Model with random, smoothly time-varying parameters .....	208
4.4 Discriminative learning of production models' parameters .....	210
5. Other types of computational models of speech production .....	210
6. Summary and discussions .....	212
References .....	212

### **Articulatory Features and Associated Production Models in Statistical Speech Recognition**

Li Deng .....	214
1. Introduction .....	214
2. Functional description of human speech communication as an encoding-decoding process .....	214
3. Overview of theories of speech perception .....	215
4. A general framework of statistical speech recognition .....	216
5. Brief analysis of weaknesses of current speech recognition technology .....	217
6. Phonological model: Overlapping articulatory features and related HMMs ...	218
7. Task-dynamic model of speech production .....	219
8. Interfacing overlapping features to task-dynamic model and a general architecture for speech recognition .....	220
9. Discussions: Machine speech recognition .....	220
References .....	223

### **Talker Normalization with Articulatory Analysis-by-Synthesis**

Richard S. McGowan .....	225
1. Introduction .....	225
2. Normalization Procedure .....	226
3. Experiments .....	229
4. Conclusion .....	231
References .....	231

### **The Psycholinguistics of Spoken Word Recognition**

Cynthia M. Connine and Thomas Deelman .....	233
1. Introduction .....	233
2. Overview: Models of spoken word recognition .....	233
3. Currency of mapping: units and the nature of lexical representations .....	235
4. Temporal nature of speech: early vs delayed commitment .....	237
4.1 Delayed commitment .....	238
5. Multiple lexical hypotheses, lexical competition and graded activation .....	239
6. Language architecture: Lexical and segmental levels .....	242
7. Language architecture: Lexical and sentential .....	244
8. Contribution of attention .....	245
References .....	247

### **Issues in Using Models for Self Evaluation and Correction of Speech**

Marie-Christine Haton .....	252
1. Introduction .....	252
2. Using models .....	253
3. Norm building .....	254
4. Matching between the subject's world and the technical world .....	255
5. Settlement of the speech education program .....	256



6. Management of the education program .....	257
7. Conclusion .....	257
References .....	257

### **The Use of the Maximum Likelihood Criterion in Language Modelling**

Hermann Ney .....	259
1. Introduction .....	259
2. Perplexity and Maximum Likelihood .....	260
3. Smoothing and Discounting for Sparse Data .....	263
3.1 Model-free Discounting and Turing-Good Estimates .....	263
3.2 Absolute Discounting .....	266
4. Partitioning-Based Models .....	267
4.1 Equivalence Classes of Histories and Decision Trees .....	267
4.2 Two-Sided Partitionings and Word Classes .....	270
5. Word Trigger Pairs .....	272
6. Maximum Entropy Approach .....	275
7. Conclusions .....	277
References .....	277

### **Language Model Adaptation**

Renato DeMon and Marcello Federico .....	280
1. Introduction .....	280
2. Background on Language Models .....	281
3. Adaptation paradigms .....	283
3.1 LM adaptation in dialogue systems .....	284
4. Basic statistical methods .....	285
4.1 Maximum a-posteriori estimation .....	285
4.2 Linear interpolation .....	286
4.3 Sublanguages mixture adaptation .....	288
4.4 Backing-off .....	288
4.5 Maximum Entropy .....	290
4.6 Minimum Discrimination Information .....	291
4.7 Generalized iterative scaling .....	292
4.8 Cache model and word triggers .....	293
5. Practical applications of adaptation paradigms .....	295
5.1 The 1993 ARPA evaluation method .....	295
5.2 Mixture based adaptation .....	296
5.3 Adaptation with a cache model .....	298
5.4 ME and MDI adaptation .....	299
5.5 LM adaptation in interactive systems .....	299
6. Conclusion .....	301
References .....	301

**Using Natural-Language Knowledge Sources in Speech Recognition**

Robert C. Moore .....	304
1. Introduction .....	304
2. Issues in Language Modeling for Speech Recognition .....	305
3. Formal Models for Natural Language .....	307
3.1 Finite-State Grammars .....	307
3.2 Context-Free Grammars .....	308
3.3 Augmented Context-Free Grammars .....	309
3.4 Expressive Power of Grammar Formalisms and the Requirements of Natural Language .....	310
4. Search Architectures for Natural-Language-Based Language Models .....	312
4.1 Word Lattice Parsing .....	312
4.2 <i>N</i> -best Filtering or Rescoring .....	312
4.3 Dynamic Generation of Partial Grammar Networks .....	313
5. Compiling Unification Grammars into Context-Free Grammars .....	314
5.1 Instantiating Unification Grammars .....	314
5.2 Removing Left Recursion from Context-Free Grammars .....	316
6. Robust Natural-Language-Based Language Models .....	318
6.1 Combining Linguistics and Statistics in a Language Model .....	318
6.2 FuHy Statistical Natural-Language Grammars .....	320
7. Summary .....	325
References .....	326

**How May I Help You?**

A.L. Gorin, G. Riccardi and J.R. Wright .....	328
1. Introduction .....	328
2. A Spoken Dialog System .....	329
3. Database .....	331
4. Algorithms .....	333
4.1 Salient Fragment Acquisition .....	335
4.2 Recognizing Fragments in Speech .....	339
4.3 Call Classification .....	341
5. Experiment Results .....	343
6. Conclusions .....	347
References .....	348

**Introduction of Rules into a Stochastic Approach for Language Modelling**

Thierry Spriet and Marc El-Bèze .....	350
1. Introduction .....	350
2. Stack Decoding Strategy .....	351
2.1 The Algorithm .....	351
2.2 The Evaluation Function .....	351
2.3 Peculiar Advantages of the Algorithm .....	352

3. Rules .....	353
3.1 Correction of Biases .....	353
3.2 Under-represented Structures and Long Span Dependencies .....	353
4. Multi Level Interactions .....	354
4.1 Linguistic and Syntactic .....	354
4.2 Phonology .....	354
5. Conclusion .....	355
References .....	355

### **History Integration into Semantic Classification**

Mauro Cettolo and Anna Corazza .....	356
1. Introduction .....	356
2. Classifier .....	357
3. Data .....	357
4. Dialogue History Integration .....	358
5. Discussion .....	360
References .....	361

### **Multilingual Speech Recognition**

E. Nöth and S. Harbeck and H. Niemann .....	362
1. Introduction .....	362
2. Architecture of the National SQEL Demonstrators .....	363
3. Language Identification with Different Amounts of Knowledge about the Training Data .....	364
3.1 A System with Explicit Language Identification .....	365
3.2 A System with Implicit Language Identification .....	367
3.3 Language Identification Based on Cepstral Feature Vectors .....	369
4. Results .....	370
5. Conclusions and Future Work .....	373
References .....	373

### **Toward ALISP: A proposal for Automatic Language Independent Speech Processing.**

Gerard Chollet, Jan Černocký, Andrei Constantinescu, Sabine Deligne and Frédéric Bimbot .....	375
1. Introduction .....	375
2. Practical benefit of ALISP .....	376
3. Issues specific to ALISP .....	377
3.1 Selecting features .....	377
3.2 Modeling speech units .....	377
3.3 Defining a derivation criterion .....	378
3.4 Building a lexicon .....	378
4. Some tools for ALISP .....	379
4.1 Temporal Decomposition .....	379

4.2 The multigram model .....	380
5. Experiments .....	381
5.1 Cross-Language Recognition .....	381
5.2 Very low bit rate speech coding .....	382
5.3 Mono-Speaker Continuous Speech Recognition .....	384
6. Conclusions .....	386
References .....	387

### **Interactive Translation of Conversational Speech**

Alex Waibel .....	389
1. Introduction .....	389
2. Background .....	390
2.1 The Problem of Spoken Language Translation .....	390
2.2 Research Efforts on Speech Translation .....	391
3. JANUS-II - A Conversational Speech Translator .....	392
3.1 Task Domains and Data Collection .....	392
3.2 System Description .....	394
3.3 Performance Evaluation .....	398
4. Applications and Forms of Deployment .....	400
4.1 Interactive Dialog Translation .....	401
4.2 Portable Speech Translation Device .....	402
4.3 Passive Simultaneous Dialog Translation .....	402
References .....	403

### **Multimodal Speech Systems**

Françoise D. Néel and Wolfgang M. Minker .....	404
1. Introduction .....	404
2. System Architecture: Knowledge Sources and Controllers .....	405
2.1 Environnient Model .....	406
2.2 System Model .....	406
2.3 User Model .....	408
2.4 Task Model .....	410
2.5 Dialogue Model .....	411
2.6 Models Interdependency .....	413
2.7 Role of Speech in Multimodal Applications .....	413
3. Information Speech Systems .....	414
3.1 Spontaneous Language Characteristics .....	414
3.2 Case Grammar Formalism used for Task Modelling .....	416
3.3 Different Parsing Methods .....	417
3.4 Task and Dialogue Model Integration .....	426
4. Conclusion .....	427
References .....	428

**Multimodal Interfaces for Multimedia Information Agents**

Alex Waibel and Bernhard Suhm and Minh Tue Vo and Jie Yang .....	431
1. Introduction .....	431
2. Interpretation of Multimodal Input .....	432
2.1 Multimodal Components .....	432
2.2 Joint Interpretation .....	432
3. Multimodal Error Correction .....	433
3.1 Multimodal Interactive Error Repair .....	433
3.2 Error Repair for Multimedia Information Agents .....	433
3.3 Evaluating Interactive Error Repair .....	434
4. Multimodal Information Agents .....	434
4.1 Information Access .....	434
4.2 Information Creation .....	435
4.3 Information Manipulation .....	435
4.4 Information Dissemination .....	436
4.5 Controlling the Interface .....	436
5. The QuickDoc Application .....	437
6. Conclusions .....	437
References .....	438
<b>Index</b> .....	440

Computational Models of Speech Pattern Processing

Ponting, K. (Ed.)

1999, XXX, 446 p., Hardcover

ISBN: 978-3-540-65478-0