

Contents

1	Introduction	1
1.1	Automatic Speech Understanding (ASU)	1
1.1.1	Applications	1
1.1.2	Application Specific Requirements	3
1.1.3	Basic ASU System Structure	4
1.2	Prosody in ASU	8
1.2.1	Potentials	8
1.2.2	General Difficulties	10
1.2.3	State-of-the-art	11
1.3	About this Book	13
1.3.1	Prerequisites and General Remarks	13
1.3.2	Contribution to the Progress in Science	15
1.3.3	Overview	18
2	Basic Approaches	21
2.1	General Considerations	21
2.2	Neural Networks (NNs)	26
2.2.1	Definitions and Computation	26
2.2.2	Training	28
2.2.3	Properties of Feed-forward NNs	30
2.2.4	Extensions to Feed-forward NNs	30
2.3	Hidden Markov Models (HMMs)	32
2.3.1	Structure	32
2.3.2	Speech Recognition with HMMs	34
2.3.3	Parameter Estimation	37
2.4	n -gram Grammars	40
2.5	Viterbi Search	43
2.6	A* Search	44
2.7	NN/HMM Hybrid Systems	47
2.7.1	NN for Feature Transformation: NN-Feat/HMM	48

2.7.2	NN as Probability Estimator: NN-Prob/HMM	52
2.8	Multi-level Semantic Classification Trees (MSCTs)	54
2.8.1	Semantic Classification Trees	55
2.8.2	Multi-level Information	56
2.8.3	The Training of MSCTs	60
2.9	Summary	63
3	The Baseline ASU Systems	67
3.1	Speech Recognition	67
3.2	Interfaces between ASR and Linguistic Analysis	71
3.3	Information Retrieval: EVAR	74
3.4	Speech-to-speech Translation: VERBMOBIL	80
3.5	Summary	92
4	Prosody	95
4.1	Definition of Basic Terms	95
4.1.1	Basic Prosodic Attributes	95
4.1.2	Compound Prosodic Attributes	99
4.2	The Functional Roles of Prosody in Human-human Communication	110
4.2.1	Remarks on the Importance of Prosody	110
4.2.2	Sentence Mood	111
4.2.3	Phrase Structure	113
4.2.4	Focus and Scope	118
4.2.5	Pragmatics and Discourse	121
4.2.6	Contrastive Accent	123
4.2.7	Emphatic Accent and other Paralinguistic Roles	123
4.2.8	Disfluencies	124
4.3	Prosody in Man-machine Communication: A Literature Survey	126
4.3.1	Word recognition	126
4.3.2	Linguistic Analysis	128
4.3.3	Stochastic Language Models	129
4.3.4	Other fields of ASU	130
4.4	Summary	131
5	Prosodic Labeling and Speech Corpora	135
5.1	ERBA	136
5.1.1	The Speech Corpus	136
5.1.2	Automatic Labeling of Syntactic-prosodic Phrase Boundaries	141

5.1.3	Automatic Labeling of Syntactic-prosodic Accents	144
5.1.4	Automatic Labels versus Listener Judgments	148
5.2	VERBMOBIL	152
5.2.1	The Speech Corpus	153
5.2.2	Acoustic-prosodic Boundary (B) and Accent Labels	158
5.2.3	Tone Labels	160
5.2.4	The Syntactic S Labels	162
5.2.5	The Syntactic-prosodic M Labels	164
5.2.6	Dialog Act Boundaries: the D Labels	174
5.2.7	Correspondence between the Boundary Labels	176
5.2.8	Correspondence of the Different Boundary Labels with Pauses	182
5.3	Time-of-day Expressions	182
5.4	The Munich SPONTAN Data	184
5.5	Summary	184
6	Preprocessing and Classification	187
6.1	Neural Inverse Filtering	187
6.2	DPF0-SEQ: Incremental Fundamental Frequency Determination	190
6.3	Classification of Sentence Mood	193
6.4	Recognition of Phrase Boundaries and Accents	196
6.4.1	General Considerations	196
6.4.2	Feature Extraction	197
6.4.3	Classification	200
6.4.4	Considerations Regarding the Integration of NNs in ASU	205
6.5	Summary	207
7	Prosodic Phrase Models	211
7.1	NN/HMM Hybrids	212
7.2	Polygram Classifier and Combination with NN	220
7.3	An Integrated Speech and Language Model using MSCTs	229
7.4	Summary	234
8	Integration of the Prosodic Attributes in ASU	237
8.1	Prosodic Scoring of Word Hypotheses Graphs	238
8.2	Word Recognition	246
8.3	Syntax	248
8.3.1	General Considerations	248
8.3.2	Parsing of Prosodically Scored Word Graphs	255
8.3.3	Constraining the Position of Head Traces	269

8.4	Semantics	272
8.5	Transfer	275
8.6	Dialog	276
8.6.1	Classification of Dialog Acts in VERBMOBIL	276
8.6.2	Prosodic Dialog Control in EVAR	278
8.7	Summary	281
9	Future Work	285
10	Summary	291
	Bibliography	299
	Glossary	335
	Index	341
A	Equations and Tables	347
A.1	Energy Computation	347
A.2	Intrinsic Normalization of the Phone Duration	348
A.3	Phonetic Alphabets	349
A.4	Examples for Feature Sets for the Acoustic-prosodic Classification	350
B	Examples	355
B.1	Ja zur Not geht's auch am Samstag	355
B.2	Speech Signals on the WORLD WIDE WEB	355
B.3	A Prosodically Classified Word Graph	357