

# Contents

<b>Forward</b>	<b>xi</b>
<b>Preface</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction	1
1.2 RSR over Digital Channels	4
1.3 Organization of the Book	5
<b>2 Speech Recognition with HMMs</b>	<b>7</b>
2.1 Introduction	7
2.2 Some General Issues	8
2.3 Analysis of Speech Signals	10
2.3.1 <i>Preprocessing of the Speech Signal</i>	10
2.3.2 <i>Linear Prediction Analysis</i>	11
2.3.3 <i>Mel-Frequency Filterbanks</i>	13
2.3.4 <i>Cepstral Coefficients</i>	13
2.3.5 <i>Other Features</i>	15
2.4 Vector Quantization	16
2.5 Approaches to ASR	18
2.5.1 <i>Pattern Matching</i>	18
2.5.2 <i>The Statistical Approach</i>	19
2.6 Hidden Markov Models	20
2.6.1 <i>Markov Processes and Hidden Markov Processes</i>	20
2.6.2 <i>Definition of a Discrete HMM</i>	22
2.6.3 <i>The Three Basic Problems</i>	23
2.6.4 <i>Generalization and Types of HMM Modeling</i>	28
2.6.5 <i>Simplifications to Continuous HMM Modeling</i>	28
2.7 Application of HMMs to Speech Recognition	29
2.8 Model Adaptation	32
2.8.1 <i>Maximum Likelihood Linear Regression (MLLR)</i>	33
2.8.2 <i>Maximum a Posteriori Linear Regression (MAPLR)</i>	35
2.9 Dealing with Uncertainty	36
2.9.1 <i>Exponential Weighting</i>	36
2.9.2 <i>Bayesian Optimal Classification with Uncertain Data</i>	37

<b>3</b>	<b>Networks and Degradation</b>	<b>41</b>
3.1	Introduction	41
3.2	Mobile and Wireless Networks	41
3.2.1	<i>Cellular Structure of a Mobile Network</i>	43
3.2.2	<i>Example of a Mobile Network Architecture: GSM/GPRS</i>	43
3.2.3	<i>Degradation in Wireless Networks</i>	46
3.2.4	<i>Wireless Channel Models for RSR</i>	50
3.2.5	<i>Implementation of RSR Systems over Mobile Networks</i>	54
3.3	IP Networks	54
3.3.1	<i>The TCP/IP Protocol Suite</i>	55
3.3.2	<i>Degradation in IP Networks</i>	58
3.3.3	<i>Lossy Packet Channel Models</i>	60
3.3.4	<i>Implementation of RSR Systems over Packet Networks</i>	65
3.4	The Acoustic Environment	66
3.4.1	<i>Additive Noise</i>	67
3.4.2	<i>Channel Distortion</i>	72
3.4.3	<i>A Model of Environment</i>	74
3.4.4	<i>Probability Distributions of Noisy Speech Features</i>	77
<b>4</b>	<b>Speech Compression and Architectures for RSR</b>	<b>85</b>
4.1	Introduction	85
4.2	Speech Coding	86
4.2.1	<i>Waveform Coders</i>	86
4.2.2	<i>Parametric Coders</i>	91
4.2.3	<i>Pitch Estimation</i>	92
4.2.4	<i>Hybrid Coders</i>	93
4.3	Recognition from Decoded Speech	100
4.3.1	<i>Effect of Coding</i>	100
4.3.2	<i>Effect of Tandeming</i>	102
4.3.3	<i>Treatment of the "Coding" Degradation</i>	103
4.4	Recognition from Codec Parameters	105
4.4.1	<i>Robustness of Bitstream-based NSR</i>	106
4.4.2	<i>Log-Energy Computation in B-NSR Systems</i>	111
4.4.3	<i>Alternative Parameter Conversion</i>	111
4.5	Distributed Speech Recognition	112
4.5.1	<i>Scalar Quantizers</i>	113
4.5.2	<i>Vector Quantizers and Product Codes</i>	113
4.5.3	<i>Very Low Bitrate PLP-based Compression for DSR</i>	120
4.5.4	<i>Transform Coders</i>	121
4.5.5	<i>Scalability in DSR Systems</i>	126
4.6	Comparison between NSR and DSR	127
<b>5</b>	<b>Robustness Against Transmission Channel Errors</b>	<b>131</b>
5.1	Introduction	131
5.2	Channel Coding Techniques	133
5.2.1	<i>Error Detection</i>	133
5.2.2	<i>Error Correction and Unequal Error Protection</i>	135
5.2.3	<i>Example of Channel Coding for NSR: GSM-EFR</i>	138
5.2.4	<i>Frame-Level Interleaving</i>	139
5.2.5	<i>Media-specific FEC</i>	140

---

5.3	Error Concealment (EC)	141
5.3.1	<i>Interpolation</i>	142
5.3.2	<i>Estimation</i>	145
5.3.3	<i>Recognizer-based EC Techniques</i>	151
5.3.4	<i>Error Concealment for NSR</i>	156
<b>6</b>	<b>Front-end Processing for Robust Feature Extraction</b>	<b>159</b>
6.1	Introduction	159
6.2	Noise Reduction Techniques	160
6.2.1	<i>Wiener Filters</i>	161
6.2.2	<i>Short-time Spectral Attenuation</i>	167
6.3	Voice Activity Detection	174
6.3.1	<i>Full-band-energy-based VAD</i>	175
6.3.2	<i>Statistical VAD</i>	177
6.3.3	<i>Using Long-term Information</i>	178
6.4	Feature Normalization	182
6.4.1	<i>Cepstral Mean Normalization</i>	182
6.4.2	<i>Frequency Analysis of Time-filtered Features</i>	185
6.4.3	<i>RASTA Processing</i>	187
6.4.4	<i>Cepstral Mean and Variance Normalization</i>	187
6.4.5	<i>Nonlinear Feature Normalization</i>	189
<b>7</b>	<b>Standards for Distributed Speech Recognition</b>	<b>197</b>
7.1	Introduction	197
7.2	Signal Preprocessing	199
7.2.1	<i>Two-stage Mel-warped Wiener Filtering</i>	199
7.2.2	<i>Offset Compensation and Waveform Processing</i>	206
7.3	Feature Extraction	207
7.3.1	<i>Computation of the Basic Features</i>	207
7.3.2	<i>AFE/XAFE Sampling Frequency Extension to 16 kHz</i>	208
7.3.3	<i>Blind Equalization</i>	211
7.3.4	<i>Voice Activity Detection</i>	212
7.3.5	<i>Pitch and Class Estimation</i>	214
7.4	Feature Compression and Encoding	217
7.4.1	<i>Basic Feature Vector Quantization</i>	217
7.4.2	<i>Extension Features Quantization and Encoding</i>	218
7.4.3	<i>Bitstream/Payload Format and Error Protection</i>	219
7.5	Feature Decoding and Postprocessing	221
7.5.1	<i>Bitstream Decoding and Decompression</i>	221
7.5.2	<i>Error Detection and Concealment</i>	222
7.5.3	<i>Server Feature Processing</i>	223
7.5.4	<i>Pitch Tracking and Smoothing</i>	223
<b>A</b>	<b>Alternative Representations of the LPC Coefficients</b>	<b>225</b>
<b>B</b>	<b>Basic Digital Modulation Concepts</b>	<b>227</b>
<b>C</b>	<b>Review of Channel Coding Techniques</b>	<b>229</b>
C.1	Media-independent FEC	229
C.1.1	<i>Linear Block Codes</i>	229

---

<i>C.1.2 Cyclic Codes</i>	231
<i>C.1.3 Convolutional Codes</i>	232
C.2 Interleaving	234
<b>Bibliography</b>	<b>237</b>
<b>List of Acronyms</b>	<b>249</b>
<b>Index</b>	<b>253</b>