

Contents

1	Introduction	1
1.1	Objectives of Speech Coding	1
1.2	Perceptual Quality	3
1.3	Speech Signals	4
1.4	Perceptual and Source Models	5
	References	8
 Part I Basic Properties of Speech Signals		
2	Speech Production and Modelling	11
2.1	Introduction	11
2.2	Physiology and Articulation	11
2.3	Phonemes	16
2.3.1	Vowels	16
2.3.2	Consonants	18
2.4	Intonation, Rhythm and Intensity	19
2.5	Vocal Tract Models	20
2.6	Models of the Glottal Excitation	22
2.7	Obstruent Modelling	24
2.8	Nasal Cavities	25
2.9	Lips	25
2.10	System Modelling	26
	References	29
3	Principles of Entropy Coding with Perceptual Quality Evaluation	31
3.1	Source and Entropy Coding	31
3.2	Quantisation	34

3.3	Rate-Distortion Optimisation.	36
3.4	Evaluation Models and Bitrate	37
	Appendix	39
	References.	44

Part II Core Tools

4	Spectral Envelope and Perceptual Masking Models	47
4.1	Introduction	47
4.2	Linear Predictive Coding	49
4.2.1	Definition of the Linear Predictive Model	51
4.2.2	Estimation of the Autocorrelation and Windowing	53
4.2.3	Pre-processing Tools	56
4.3	Perceptual Modelling	59
4.4	Quantisation and Coding of Linear Predictive Models	61
4.4.1	Representations	61
4.4.2	Evaluation of Quantisation Quality	64
4.4.3	Vector Quantisation and Coding	65
4.4.4	Interpolation of Envelope Models	68
4.5	Summary and Discussion	69
	Appendix	70
	References.	75
5	Windowing and the Zero Input Response	77
5.1	Introduction and Motivation	77
5.2	Filter Windowing	79
5.2.1	Modelling a Stationary Signal	80
5.2.2	Adaptive Modelling.	81
5.3	Source Model Windowing	82
5.4	Perceptual Windowing	83
5.5	Evaluation and Minimisation of Perceptual Error	84
5.6	Frames and Subframes	87
5.7	Summary	88
	Appendix	88
	References.	90
6	Fundamental Frequency	91
6.1	Source Modelling	91
6.2	Long-Term Prediction.	92
6.3	Codebook Formulation	92
6.4	Stability	93
6.5	Lag Estimation	94
6.6	Lag Quantisation.	95
	References.	96

7	Residual Coding	97
7.1	Background	97
7.2	Codebook Design	98
7.3	Algebraic Codebooks	101
7.3.1	Quantisation and Codebook Design	102
7.3.2	Codebook Search	104
7.3.3	Encoding	107
7.4	Other Codebooks	113
	References	115
8	Signal Gain and Harmonics to Noise Ratio	117
8.1	Basic Gain Quantisation	117
8.2	Harmonics to Noise Ratio	119
	Reference	120

Part III Advanced Tools and Extensions

9	Pre- and Postfiltering	123
9.1	Introduction	123
9.2	Formant Enhancement and Excitation Shaping	124
9.3	Bass Postfilter	126
9.4	Pitch Sharpening	127
9.5	Pulse Dispersion and Phase Scrambling	128
	References	129
10	Frequency Domain Coding	131
10.1	Introduction	131
10.2	Overlapping Windowing	133
10.3	Windowing with Critical Sampling	137
10.4	Time-Frequency Transforms	140
10.5	Perceptual Modelling and Quantisation	143
10.6	Entropy Coding and the Rate Loop	144
10.7	Overall Codec Structure and Integration	147
10.8	Summary	149
	References	150
11	Bandwidth Extension	151
11.1	Introduction	151
11.2	Bandwidth Extension Basics	152
11.3	Bandwidth Extension Implementation	153
11.3.1	Bandwidth Extension Encoder	153
11.3.2	Bandwidth Extension Decoder	154

11.4	Common Bandwidth Extension Methods	156
11.4.1	Spectral Band Replication	157
11.4.2	Speech Bandwidth Extension	157
11.4.3	Intelligent Gap Filling	158
11.5	Conclusion	158
	References.	159
12	Packet Loss and Concealment	161
12.1	Introduction	161
12.1.1	Error Types	162
12.1.2	Methods for Concealment of Errors in Multimedia Transmissions	163
12.2	Sender Based Concealment.	163
12.2.1	Interleaving	164
12.2.2	Forward Error Correction	165
12.2.3	Multiple Description Coding and Randomised Quantisation	168
12.3	Receiver Based Methods.	168
12.3.1	Concealment	169
12.3.2	Recovery	176
12.3.3	Jitter Buffer	178
12.4	Interactive Methods	180
12.4.1	Retransmission	181
12.4.2	Adaptive or Channel-Aware Modes.	182
12.5	Conclusion	182
	References.	183
13	Voice Activity Detection	185
13.1	Introduction	185
13.1.1	Applications	185
13.1.2	Requirements.	187
13.2	Methods	188
13.2.1	Feature Extraction	188
13.2.2	Detection, Classification and Supervised Learning.	192
13.2.3	Example Implementations	195
13.3	Evaluation.	197
13.3.1	Performance Measures.	198
13.3.2	Challenges of VAD.	198
	References.	201
14	Relaxed Code-Excited Linear Prediction (RCELP).	205
14.1	Generalised Analysis-by-Synthesis	205
14.2	RCELP for Voiced Coding.	207
14.2.1	Pitch Estimation and Pulse Localisation	208
14.2.2	Continuous Linear Delay Contour.	209

14.2.3	Signal Modification	210
14.2.4	Pitch Coding and Subsequent Coding	211
14.3	Parametric Unvoiced Coding	212
14.4	RCELP in Standards	214
	References	215

Part IV Standards and Specifications

15	Quality Evaluation	219
15.1	Introduction	219
15.1.1	Applications	219
15.1.2	Aspects of Quality	220
15.1.3	Test Material	222
15.2	Subjective Evaluation	222
15.2.1	ITU-T Recommendation P.800	224
15.2.2	Recommendation ITU-T Recommendation P.835	228
15.2.3	Recommendation ITU-R BS.1534 (MUSHRA)	228
15.3	Objective Evaluation	231
15.3.1	ITU-T Recommendation P.862 (PESQ)	232
15.3.2	ITU-T Recommendation P.863 (POLQA)	233
15.3.3	ITU-R Recommendation BS.1387 (PEAQ)	234
	References	235
Index	237

Speech Coding

with Code-Excited Linear Prediction

Bäckström, T.

2017, XXI, 240 p. 76 illus., Hardcover

ISBN: 978-3-319-50202-1