

# Contents

<b>About the Authors</b>	<b>xxi</b>
<b>Other Wiley and IEEE Press Books on Related Topics</b>	<b>xxiii</b>
<b>Preface and Motivation</b>	<b>xxv</b>
<b>Acknowledgements</b>	<b>xxxv</b>
<b>I Speech Signals and Waveform Coding</b>	<b>1</b>
<b>1 Speech Signals and an Introduction to Speech Coding</b>	<b>3</b>
1.1 Motivation of Speech Compression . . . . .	3
1.2 Basic Characterisation of Speech Signals . . . . .	4
1.3 Classification of Speech Coders . . . . .	8
1.3.1 Waveform Coding . . . . .	9
1.3.1.1 Time-domain Waveform Coding . . . . .	9
1.3.1.2 Frequency-domain Waveform Coding . . . . .	10
1.3.2 Vocoders . . . . .	10
1.3.3 Hybrid Coding . . . . .	11
1.4 Waveform Coding . . . . .	11
1.4.1 Digitisation of Speech . . . . .	11
1.4.2 Quantisation Characteristics . . . . .	13
1.4.3 Quantisation Noise and Rate-distortion Theory . . . . .	14
1.4.4 Non-uniform Quantisation for a known PDF: Companding . . . . .	16
1.4.5 PDF-independent Quantisation using Logarithmic Compression . . . . .	18
1.4.5.1 The $\mu$ -law Compressor . . . . .	20
1.4.5.2 The A-law Compressor . . . . .	21
1.4.6 Optimum Non-uniform Quantisation . . . . .	23
1.5 Chapter Summary . . . . .	28

<b>2</b>	<b>Predictive Coding</b>	<b>29</b>
2.1	Forward-Predictive Coding	29
2.2	DPCM Codec Schematic	30
2.3	Predictor Design	31
2.3.1	Problem Formulation	31
2.3.2	Covariance Coefficient Computation	33
2.3.3	Predictor Coefficient Computation	34
2.4	Adaptive One-word-memory Quantisation	39
2.5	DPCM Performance	40
2.6	Backward-adaptive Prediction	42
2.6.1	Background	42
2.6.2	Stochastic Model Processes	44
2.7	The 32 kbps G.721 ADPCM Codec	47
2.7.1	Functional Description of the G.721 Codec	47
2.7.2	Adaptive Quantiser	47
2.7.3	G.721 Quantiser Scale Factor Adaptation	48
2.7.4	G.721 Adaptation Speed Control	50
2.7.5	G.721 Adaptive Prediction and Signal Reconstruction	51
2.8	Subjective and Objective Speech Quality	53
2.9	Variable-rate G.726 and Embedded G.727 ADPCM	54
2.9.1	Motivation	54
2.9.2	Embedded G.727 ADPCM Coding	55
2.9.3	Performance of the Embedded G.727 ADPCM Codec	56
2.10	Rate-distortion in Predictive Coding	62
2.11	Chapter Summary	67
<b>II</b>	<b>Analysis-by-Synthesis Coding</b>	<b>69</b>
<b>3</b>	<b>Analysis-by-Synthesis Principles</b>	<b>71</b>
3.1	Motivation	71
3.2	Analysis-by-Synthesis Codec Structure	72
3.3	The Short-term Synthesis Filter	73
3.4	Long-term Prediction	76
3.4.1	Open-loop Optimisation of LTP Parameters	76
3.4.2	Closed-loop Optimisation of LTP Parameters	80
3.5	Excitation Models	85
3.6	Adaptive Short-term and Long-term Post-Filtering	88
3.7	Lattice-based Linear Prediction	90
3.8	Chapter Summary	97
<b>4</b>	<b>Speech Spectral Quantisation</b>	<b>99</b>
4.1	Log-area Ratios	99
4.2	Line Spectral Frequencies	103
4.2.1	Derivation of the Line Spectral Frequencies	103
4.2.2	Computation of the Line Spectral Frequencies	107

4.2.3	Chebyshev Description of Line Spectral Frequencies . . . . .	109
4.3	Vector Quantisation of Spectral Parameters . . . . .	115
4.3.1	Background . . . . .	115
4.3.2	Speaker-adaptive Vector Quantisation of LSFs . . . . .	115
4.3.3	Stochastic VQ of LPC Parameters . . . . .	117
4.3.3.1	Background . . . . .	117
4.3.3.2	The Stochastic VQ Algorithm . . . . .	118
4.3.4	Robust Vector Quantisation Schemes for LSFs . . . . .	121
4.3.5	LSF VQs in Standard Codecs . . . . .	122
4.4	Spectral Quantisers for Wideband Speech Coding . . . . .	123
4.4.1	Introduction to Wideband Spectral Quantisation . . . . .	123
4.4.1.1	Statistical Properties of Wideband LSFs . . . . .	125
4.4.1.2	Speech Codec Specifications . . . . .	127
4.4.2	Wideband LSF VQs . . . . .	128
4.4.2.1	Memoryless Vector Quantisation . . . . .	128
4.4.2.2	Predictive Vector Quantisation . . . . .	132
4.4.2.3	Multimode Vector Quantisation . . . . .	133
4.4.3	Simulation Results and Subjective Evaluations . . . . .	136
4.4.4	Conclusions on Wideband Spectral Quantisation . . . . .	137
4.5	Chapter Summary . . . . .	138
<b>5</b>	<b>Regular Pulse Excited Coding</b> . . . . .	<b>139</b>
5.1	Theoretical Background . . . . .	139
5.2	The 13 kbps RPE-LTP GSM Speech Encoder . . . . .	146
5.2.1	Pre-processing . . . . .	146
5.2.2	STP Analysis Filtering . . . . .	148
5.2.3	LTP Analysis Filtering . . . . .	148
5.2.4	Regular Excitation Pulse Computation . . . . .	149
5.3	The 13 kbps RPE-LTP GSM Speech Decoder . . . . .	151
5.4	Bit-sensitivity of the 13 kbps GSM RPE-LTP Codec . . . . .	153
5.5	Application Example: A Tool-box Based Speech Transceiver . . . . .	154
5.6	Chapter Summary . . . . .	157
<b>6</b>	<b>Forward-Adaptive Code Excited Linear Prediction</b> . . . . .	<b>159</b>
6.1	Background . . . . .	159
6.2	The Original CELP Approach . . . . .	160
6.3	Fixed Codebook Search . . . . .	163
6.4	CELP Excitation Models . . . . .	165
6.4.1	Binary-pulse Excitation . . . . .	165
6.4.2	Transformed Binary-pulse Excitation . . . . .	166
6.4.2.1	Excitation Generation . . . . .	166
6.4.2.2	Bit-sensitivity Analysis of the 4.8 Kbps TBPE Speech Codec . . . . .	168
6.4.3	Dual-rate Algebraic CELP Coding . . . . .	170
6.4.3.1	ACELP Codebook Structure . . . . .	170
6.4.3.2	Dual-rate ACELP Bit Allocation . . . . .	172

	6.4.3.3	Dual-rate ACELP Codec Performance . . . . .	173
6.5		Optimisation of the CELP Codec Parameters . . . . .	174
	6.5.1	Introduction . . . . .	174
	6.5.2	Calculation of the Excitation Parameters . . . . .	175
	6.5.2.1	Full Codebook Search Theory . . . . .	175
	6.5.2.2	Sequential Search Procedure . . . . .	177
	6.5.2.3	Full Search Procedure . . . . .	178
	6.5.2.4	Sub-optimal Search Procedures . . . . .	180
	6.5.2.5	Quantisation of the Codebook Gains . . . . .	181
	6.5.3	Calculation of the Synthesis Filter Parameters . . . . .	183
	6.5.3.1	Bandwidth Expansion . . . . .	184
	6.5.3.2	Least Squares Techniques . . . . .	184
	6.5.3.3	Optimisation via Powell's Method . . . . .	187
	6.5.3.4	Simulated Annealing and the Effects of Quantisation . . . . .	188
6.6		The Error Sensitivity of CELP Codecs . . . . .	192
	6.6.1	Introduction . . . . .	192
	6.6.2	Improving the Spectral Information Error Sensitivity . . . . .	192
	6.6.2.1	LSF Ordering Policies . . . . .	192
	6.6.2.2	The Effect of FEC on the Spectral Parameters . . . . .	195
	6.6.2.3	The Effect of Interpolation . . . . .	195
	6.6.3	Improving the Error Sensitivity of the Excitation Parameters . . . . .	196
	6.6.3.1	The Fixed Codebook Index . . . . .	197
	6.6.3.2	The Fixed Codebook Gain . . . . .	197
	6.6.3.3	Adaptive Codebook Delay . . . . .	198
	6.6.3.4	Adaptive Codebook Gain . . . . .	199
	6.6.4	Matching Channel Codecs to the Speech Codec . . . . .	199
	6.6.5	Error Resilience Conclusions . . . . .	203
6.7		Application Example: A Dual-mode 3.1 kBd Speech Transceiver . . . . .	204
	6.7.1	The Transceiver Scheme . . . . .	204
	6.7.2	Re-configurable Modulation . . . . .	205
	6.7.3	Source-matched Error Protection . . . . .	206
	6.7.3.1	Low-quality 3.1 kBd Mode . . . . .	206
	6.7.3.2	High-quality 3.1 kBd Mode . . . . .	210
	6.7.4	Voice Activity Detection and Packet Reservation Multiple Access . . . . .	211
	6.7.5	3.1 kBd System Performance . . . . .	214
	6.7.6	3.1 kBd System Summary . . . . .	217
6.8		Multi-slot PRMA Transceiver . . . . .	218
	6.8.1	Background and Motivation . . . . .	218
	6.8.2	PRMA-assisted Multi-slot Adaptive Modulation . . . . .	219
	6.8.3	Adaptive GSM-like Schemes . . . . .	220
	6.8.4	Adaptive DECT-like Schemes . . . . .	222
	6.8.5	Summary of Adaptive Multi-slot PRMA . . . . .	223
6.9		Chapter Summary . . . . .	223

<b>7</b>	<b>Standard Speech Codecs</b>	<b>225</b>
7.1	Background . . . . .	225
7.2	The US DoD FS-1016 4.8 kbps CELP Codec . . . . .	225
7.2.1	Introduction . . . . .	225
7.2.2	LPC Analysis and Quantisation . . . . .	227
7.2.3	The Adaptive Codebook . . . . .	228
7.2.4	The Fixed Codebook . . . . .	229
7.2.5	Error Concealment Techniques . . . . .	230
7.2.6	Decoder Post-filtering . . . . .	231
7.2.7	Conclusion . . . . .	231
7.3	The 7.95 kbps Pan-American Speech Codec – Known as IS-54 DAMPS Codec . . . . .	231
7.4	The 6.7 kbps Japanese Digital Cellular System’s Speech Codec . . . . .	235
7.5	The Qualcomm Variable Rate CELP Codec . . . . .	237
7.5.1	Introduction . . . . .	237
7.5.2	Codec Schematic and Bit Allocation . . . . .	238
7.5.3	Codec Rate Selection . . . . .	239
7.5.4	LPC Analysis and Quantisation . . . . .	240
7.5.5	The Pitch Filter . . . . .	241
7.5.6	The Fixed Codebook . . . . .	242
7.5.7	Rate 1/8 Filter Excitation . . . . .	243
7.5.8	Decoder Post-filtering . . . . .	243
7.5.9	Error Protection and Concealment Techniques . . . . .	244
7.5.10	Conclusion . . . . .	244
7.6	Japanese Half-rate Speech Codec . . . . .	245
7.6.1	Introduction . . . . .	245
7.6.2	Codec Schematic and Bit Allocation . . . . .	245
7.6.3	Encoder Pre-processing . . . . .	247
7.6.4	LPC Analysis and Quantisation . . . . .	248
7.6.5	The Weighting Filter . . . . .	248
7.6.6	Excitation Vector 1 . . . . .	249
7.6.7	Excitation Vector 2 . . . . .	250
7.6.8	Channel Coding . . . . .	251
7.6.9	Decoder Post-processing . . . . .	252
7.7	The Half-rate GSM Speech Codec . . . . .	253
7.7.1	Half-rate GSM Codec Outline and Bit Allocation . . . . .	253
7.7.2	Spectral Quantisation in the Half-rate GSM Codec . . . . .	255
7.7.3	Error Protection . . . . .	256
7.8	The 8 kbps G.729 Codec . . . . .	257
7.8.1	Introduction . . . . .	257
7.8.2	Codec Schematic and Bit Allocation . . . . .	257
7.8.3	Encoder Pre-processing . . . . .	258
7.8.4	LPC Analysis and Quantisation . . . . .	259
7.8.5	The Weighting Filter . . . . .	262
7.8.6	The Adaptive Codebook . . . . .	262
7.8.7	The Fixed Algebraic Codebook . . . . .	263

7.8.8	Quantisation of the Gains . . . . .	266
7.8.9	Decoder Post-processing . . . . .	267
7.8.10	G.729 Error-concealment Techniques . . . . .	269
7.8.11	G.729 Bit-sensitivity . . . . .	270
7.8.12	Turbo-coded Orthogonal Frequency Division Multiplex Transmission of G.729 Encoded Speech . . . . .	271
7.8.12.1	Background . . . . .	271
7.8.12.2	System Overview . . . . .	272
7.8.12.3	Turbo Channel Encoding . . . . .	273
7.8.12.4	OFDM in the FRAMES Speech/Data Sub-burst . . . . .	274
7.8.12.5	Channel Model . . . . .	275
7.8.12.6	Turbo-coded G.729 OFDM Parameters . . . . .	275
7.8.12.7	Turbo-coded G.729 OFDM Performance . . . . .	276
7.8.12.8	Turbo-coded G.729 OFDM Summary . . . . .	277
7.8.13	G.729 Summary . . . . .	278
7.9	The Reduced Complexity G.729 Annex A Codec . . . . .	278
7.9.1	Introduction . . . . .	278
7.9.2	The Perceptual Weighting Filter . . . . .	279
7.9.3	The Open-loop Pitch Search . . . . .	280
7.9.4	The Closed-loop Pitch Search . . . . .	280
7.9.5	The Algebraic Codebook Search . . . . .	280
7.9.6	The Decoder Post-processing . . . . .	281
7.9.7	Conclusions . . . . .	281
7.10	The 12.2 kbps Enhanced Full-rate GSM Speech Codec . . . . .	282
7.10.1	Enhanced Full-rate GSM Codec Outline . . . . .	282
7.10.2	Enhanced Full-rate GSM Encoder . . . . .	284
7.10.2.1	Spectral Quantisation and Windowing in the Enhanced Full-rate GSM Codec . . . . .	284
7.10.2.2	Adaptive Codebook Search . . . . .	286
7.10.2.3	Fixed Codebook Search . . . . .	286
7.11	The Enhanced Full-rate 7.4 kbps IS-136 Speech Codec . . . . .	287
7.11.1	IS-136 Codec Outline . . . . .	287
7.11.2	IS-136 Bit-allocation Scheme . . . . .	289
7.11.3	Fixed Codebook Search . . . . .	290
7.11.4	IS-136 Channel Coding . . . . .	291
7.12	The ITU G.723.1 Dual-rate Codec . . . . .	292
7.12.1	Introduction . . . . .	292
7.12.2	G.723.1 Encoding Principle . . . . .	292
7.12.3	Vector-quantisation of the LSPs . . . . .	294
7.12.4	Formant-based Weighting Filter . . . . .	295
7.12.5	The 6.3 kbps High-rate G.723.1 Excitation . . . . .	296
7.12.6	The 5.3 kbps Low-rate G.723.1 Excitation . . . . .	297
7.12.7	G.723.1 Bit Allocation . . . . .	298
7.12.8	G.723.1 Error Sensitivity . . . . .	300
7.13	Advanced Multirate JD-CDMA Transceiver . . . . .	302
7.13.1	Multirate Codecs and Systems . . . . .	302

7.13.2	System Overview . . . . .	305
7.13.3	The Adaptive Multirate Speech Codec . . . . .	306
7.13.3.1	AMR Codec Overview . . . . .	306
7.13.3.2	Linear Prediction Analysis . . . . .	307
7.13.3.3	LSF Quantisation . . . . .	308
7.13.3.4	Pitch Analysis . . . . .	308
7.13.3.5	Fixed Codebook with Algebraic Structure . . . . .	308
7.13.3.6	Post-processing . . . . .	310
7.13.3.7	The AMR Codec's Bit Allocation . . . . .	311
7.13.3.8	Codec Mode Switching Philosophy . . . . .	311
7.13.4	The AMR Speech Codec's Error Sensitivity . . . . .	312
7.13.5	RRNS-based Channel Coding . . . . .	315
7.13.5.1	RRNS Overview . . . . .	315
7.13.5.2	Source-matched Error Protection . . . . .	316
7.13.6	Joint Detection Code Division Multiple Access . . . . .	318
7.13.6.1	Overview . . . . .	318
7.13.6.2	Joint Detection Based Adaptive Code Division Multiple Access . . . . .	319
7.13.7	System Performance . . . . .	319
7.13.7.1	Subjective Testing . . . . .	326
7.13.8	Conclusions . . . . .	327
7.14	Chapter Summary . . . . .	327
<b>8</b>	<b>Backward-adaptive Code Excited Linear Prediction</b>	<b>331</b>
8.1	Introduction . . . . .	331
8.2	Motivation and Background . . . . .	331
8.3	Backward-adaptive G728 Codec Schematic . . . . .	334
8.4	Backward-adaptive G728 Coding Algorithm . . . . .	336
8.4.1	G728 Error Weighting . . . . .	336
8.4.2	G728 Windowing . . . . .	337
8.4.3	Codebook Gain Adaption . . . . .	341
8.4.4	G728 Codebook Search . . . . .	343
8.4.5	G728 Excitation Vector Quantisation . . . . .	345
8.4.6	G728 Adaptive Post-filtering . . . . .	347
8.4.6.1	Adaptive Long-term Post-filtering . . . . .	348
8.4.6.2	G.728 Adaptive Short-term Post-filtering . . . . .	350
8.4.7	Complexity and Performance of the G728 Codec . . . . .	351
8.5	Reduced-rate G728-like Codec: Variable-length Excitation Vector . . . . .	351
8.6	The Effects of Long-term Prediction . . . . .	354
8.7	Closed-loop Codebook Training . . . . .	359
8.8	Reduced-rate G728-like Codec: Constant-length Excitation Vector . . . . .	364
8.9	Programmable-rate 8–4 kbps Low-delay CELP Codecs . . . . .	365
8.9.1	Motivation . . . . .	365
8.9.2	8–4 kbps Codec Improvements Due to Increasing Codebook Sizes . . . . .	366
8.9.3	8–4 kbps Codecs – Forward Adaption of the Short-term Synthesis Filter . . . . .	367

8.9.4	Forward Adaption of the Long-term Predictor . . . . .	368
8.9.4.1	Initial Experiments . . . . .	368
8.9.4.2	Quantisation of Jointly Optimized Gains . . . . .	370
8.9.4.3	8–4 kbps Codecs – Voiced/Unvoiced Codebooks . . . . .	373
8.9.5	Low-delay Codecs at 4–8 kbps . . . . .	375
8.9.6	Low-delay ACELP Codec . . . . .	378
8.10	Backward-adaptive Error Sensitivity Issues . . . . .	381
8.10.1	The Error Sensitivity of the G728 Codec . . . . .	381
8.10.2	The Error Sensitivity of our 4–8 kbps Low-delay Codecs . . . . .	382
8.10.3	The Error Sensitivity of our Low-delay ACELP Codec . . . . .	387
8.11	A Low-delay Multimode Speech Transceiver . . . . .	388
8.11.1	Background . . . . .	388
8.11.2	8–16 kbps Codec Performance . . . . .	388
8.11.3	Transmission Issues . . . . .	389
8.11.3.1	Higher-quality Mode . . . . .	389
8.11.3.2	Lower-quality Mode . . . . .	391
8.11.4	Speech Transceiver Performance . . . . .	391
8.12	Chapter Summary . . . . .	392

### III Wideband Speech, MPEG-4 Audio and Their Transmission 393

<b>9</b>	<b>Wideband Speech Coding</b>	<b>395</b>
9.1	Sub-band-ADPCM Wideband Coding at 64 kbps . . . . .	395
9.1.1	Introduction and Specifications . . . . .	395
9.1.2	G722 Codec Outline . . . . .	396
9.1.3	Principles of Sub-band Coding . . . . .	399
9.1.4	Quadrature Mirror Filtering . . . . .	400
9.1.4.1	Analysis Filtering . . . . .	400
9.1.4.2	Synthesis Filtering . . . . .	403
9.1.4.3	Practical QMF Design Constraints . . . . .	405
9.1.5	G722 Adaptive Quantisation and Prediction . . . . .	410
9.1.6	G722 Coding Performance . . . . .	412
9.2	Wideband Transform-coding at 32 kbps . . . . .	413
9.2.1	Background . . . . .	413
9.2.2	Transform-coding Algorithm . . . . .	413
9.3	Sub-band-split Wideband CELP Codecs . . . . .	416
9.3.1	Background . . . . .	416
9.3.2	Sub-band-based Wideband CELP Coding . . . . .	417
9.3.2.1	Motivation . . . . .	417
9.3.2.2	Low-band Coding . . . . .	417
9.3.2.3	High-band Coding . . . . .	418
9.3.2.4	Bit-allocation Scheme . . . . .	419
9.4	Fullband Wideband ACELP Coding . . . . .	420
9.4.1	Wideband ACELP Excitation . . . . .	420

9.4.2	Backward-adaptive 32 kbps Wideband ACELP . . . . .	422
9.4.3	Forward-adaptive 9.6 kbps Wideband ACELP . . . . .	423
9.5	A Turbo-coded Burst-by-burst Adaptive Wideband Speech Transceiver . . . . .	425
9.5.1	Background and Motivation . . . . .	425
9.5.2	System Overview . . . . .	428
9.5.3	System Parameters . . . . .	428
9.5.4	Constant Throughput Adaptive Modulation . . . . .	429
9.5.5	Adaptive Wideband Transceiver Performance . . . . .	431
9.5.6	Multi-mode Transceiver Adaptation . . . . .	432
9.5.7	Transceiver Mode Switching . . . . .	433
9.5.8	The Wideband G.722.1 Codec . . . . .	435
9.5.8.1	Audio Codec Overview . . . . .	435
9.5.9	Detailed Description of the Audio Codec . . . . .	437
9.5.10	Wideband Adaptive System Performance . . . . .	439
9.5.11	Audio Frame Error Results . . . . .	440
9.5.12	Audio SEGSNR Performance and Discussions . . . . .	441
9.5.13	G.722.1 Audio Transceiver Summary and Conclusions . . . . .	442
9.6	Turbo-detected Unequal Error Protection Irregular Convolutional Coded AMR-WB Transceivers . . . . .	442
9.6.1	Introduction . . . . .	442
9.6.2	The AMR-WB Codec's Error Sensitivity . . . . .	445
9.6.3	System Model . . . . .	445
9.6.4	Design of Irregular Convolutional Codes . . . . .	446
9.6.5	An Irregular Convolutional Code Example . . . . .	449
9.6.6	UEP AMR IRCC Performance Results . . . . .	450
9.6.7	UEP AMR Conclusions . . . . .	452
9.7	The AMR-WB+ Audio Codec . . . . .	454
9.7.1	Introduction . . . . .	454
9.7.2	Audio Requirements in Mobile Multimedia Applications . . . . .	456
9.7.2.1	Summary of Audiovisual Services . . . . .	457
9.7.2.2	Bit Rates Supported by the Radio Network . . . . .	457
9.7.3	Overview of the AMR-WB+ Codec . . . . .	459
9.7.3.1	Encoding the High Frequencies . . . . .	462
9.7.3.2	Stereo Encoding . . . . .	462
9.7.3.3	Complexity of AMR-WB+ . . . . .	463
9.7.3.4	Transport and File Format of AMR-WB+ . . . . .	463
9.7.4	Performance of AMR-WB+ . . . . .	463
9.7.5	Summary of the AMR-WB+ Codec . . . . .	465
9.8	Chapter Summary . . . . .	466
<b>10</b>	<b>MPEG-4 Audio Compression and Transmission</b>	<b>469</b>
10.1	Overview of MPEG-4 Audio . . . . .	469
10.2	General Audio Coding . . . . .	471
10.2.1	Advanced Audio Coding . . . . .	479
10.2.2	Gain Control Tool . . . . .	482
10.2.3	Psycho-acoustic Model . . . . .	482

10.2.4	Temporal Noise Shaping . . . . .	484
10.2.5	Stereophonic Coding . . . . .	486
10.2.6	AAC Quantisation and Coding . . . . .	487
10.2.7	Noiseless Huffman Coding . . . . .	489
10.2.8	Bit-sliced Arithmetic Coding . . . . .	490
10.2.9	Transform-domain Weighted Interleaved Vector Quantisation . . . . .	492
10.2.10	Parametric Audio Coding . . . . .	495
10.3	Speech Coding in MPEG-4 Audio . . . . .	495
10.3.1	Harmonic Vector Excitation Coding . . . . .	496
10.3.2	CELP Coding in MPEG-4 . . . . .	498
10.3.3	LPC Analysis and Quantisation . . . . .	500
10.3.4	Multi Pulse and Regular Pulse Excitation . . . . .	502
10.4	MPEG-4 Codec Performance . . . . .	503
10.5	MPEG-4 Space-time Block Coded OFDM Audio Transceiver . . . . .	505
10.5.1	System Overview . . . . .	506
10.5.2	System Parameters . . . . .	507
10.5.3	Frame Dropping Procedure . . . . .	507
10.5.4	Space-time Coding . . . . .	510
10.5.5	Adaptive Modulation . . . . .	513
10.5.6	System Performance . . . . .	514
10.6	Turbo-detected Space-time Trellis Coded MPEG-4 Audio Transceivers . . . . .	516
10.6.1	Motivation and Background . . . . .	516
10.6.2	Audio Turbo Transceiver Overview . . . . .	518
10.6.3	The Turbo Transceiver . . . . .	519
10.6.4	Turbo Transceiver Performance Results . . . . .	521
10.6.5	MPEG-4 Turbo Transceiver Summary . . . . .	524
10.7	Turbo-detected Space-time Trellis Coded MPEG-4 Versus AMR-WB Speech Transceivers . . . . .	525
10.7.1	Motivation and Background . . . . .	525
10.7.2	The AMR-WB Codec's Error Sensitivity . . . . .	526
10.7.3	The MPEG-4 TWINVQ Codec's Error Sensitivity . . . . .	527
10.7.4	The Turbo Transceiver . . . . .	528
10.7.5	Performance Results . . . . .	531
10.7.6	AMR-WB and MPEG-4 TWINVQ Turbo Transceiver Summary . . . . .	534
10.8	Chapter Summary . . . . .	534

## **IV Very Low-rate Coding and Transmission 537**

### **11 Overview of Low-rate Speech Coding 539**

11.1	Low-bitrate Speech Coding . . . . .	539
11.1.1	AbS Coding . . . . .	542
11.1.2	Speech Coding at 2.4 kbps . . . . .	543
11.1.2.1	Background to 2.4 kbps Speech Coding . . . . .	544
11.1.2.2	Frequency Selective Harmonic Coder . . . . .	545
11.1.2.3	Sinusoidal Transform Coder . . . . .	546

11.1.2.4	Multiband Excitation Coders . . . . .	547
11.1.2.5	Sub-band Linear Prediction Coder . . . . .	549
11.1.2.6	Mixed Excitation Linear Prediction Coder . . . . .	549
11.1.2.7	Waveform Interpolation Coder . . . . .	551
11.1.3	Speech Coding Below 2.4 kbps . . . . .	552
11.2	Linear Predictive Coding Model . . . . .	553
11.2.1	Short-term Prediction . . . . .	554
11.2.2	Long-term Prediction . . . . .	556
11.2.3	Final Analysis-by-Synthesis Model . . . . .	556
11.3	Speech Quality Measurements . . . . .	557
11.3.1	Objective Speech Quality Measures . . . . .	557
11.3.2	Subjective Speech Quality Measures . . . . .	558
11.3.3	2.4 kbps Selection Process . . . . .	558
11.4	Speech Database . . . . .	560
11.5	Chapter Summary . . . . .	563
<b>12</b>	<b>Linear Predictive Vocoder</b>	<b>565</b>
12.1	Overview of a Linear Predictive Vocoder . . . . .	565
12.2	Line Spectrum Frequencies Quantisation . . . . .	566
12.2.1	Line Spectrum Frequencies Scalar Quantisation . . . . .	566
12.2.2	Line Spectrum Frequencies Vector Quantisation . . . . .	568
12.3	Pitch Detection . . . . .	571
12.3.1	Voiced–Unvoiced Decision . . . . .	573
12.3.2	Oversampled Pitch Detector . . . . .	574
12.3.3	Pitch Tracking . . . . .	578
12.3.3.1	Computational Complexity . . . . .	581
12.3.4	Integer Pitch Detector . . . . .	582
12.4	Unvoiced Frames . . . . .	583
12.5	Voiced Frames . . . . .	584
12.5.1	Placement of Excitation Pulses . . . . .	585
12.5.2	Pulse Energy . . . . .	585
12.6	Adaptive Postfilter . . . . .	585
12.7	Pulse Dispersion Filter . . . . .	588
12.7.1	Pulse Dispersion Principles . . . . .	588
12.7.2	Pitch Independent Glottal Pulse Shaping Filter . . . . .	589
12.7.3	Pitch-dependent Glottal Pulse Shaping Filter . . . . .	592
12.8	Results for Linear Predictive Vocoder . . . . .	592
12.9	Chapter Summary . . . . .	597
<b>13</b>	<b>Wavelets and Pitch Detection</b>	<b>599</b>
13.1	Conceptual Introduction to Wavelets . . . . .	599
13.1.1	Fourier Theory . . . . .	599
13.1.2	Wavelet Theory . . . . .	601
13.1.3	Detecting Discontinuities with Wavelets . . . . .	601
13.2	Introduction to Wavelet Mathematics . . . . .	602
13.2.1	Multiresolution Analysis . . . . .	603

13.2.2	Polynomial Spline Wavelets . . . . .	604
13.2.3	Pyramidal Algorithm . . . . .	605
13.2.4	Boundary Effects . . . . .	607
13.3	Preprocessing the Wavelet Transform Signal . . . . .	607
13.3.1	Spurious Pulses . . . . .	609
13.3.2	Normalisation . . . . .	610
13.3.3	Candidate Glottal Pulses . . . . .	610
13.4	Voiced–unvoiced Decision . . . . .	610
13.5	Wavelet-based Pitch Detector . . . . .	612
13.5.1	Dynamic Programming . . . . .	613
13.5.2	Autocorrelation Simplification . . . . .	616
13.6	Chapter Summary . . . . .	619
<b>14</b>	<b>Zinc Function Excitation</b>	<b>621</b>
14.1	Introduction . . . . .	621
14.2	Overview of Prototype Waveform Interpolation Zinc Function Excitation . . . . .	622
14.2.1	Coding Scenarios . . . . .	622
14.2.1.1	U–U–U Encoder Scenario . . . . .	624
14.2.1.2	U–U–V Encoder Scenario . . . . .	624
14.2.1.3	V–U–U Encoder Scenario . . . . .	625
14.2.1.4	U–V–U Encoder Scenario . . . . .	625
14.2.1.5	V–V–V Encoder Scenario . . . . .	625
14.2.1.6	V–U–V Encoder Scenario . . . . .	626
14.2.1.7	U–V–V Encoder Scenario . . . . .	626
14.2.1.8	V–V–U Encoder Scenario . . . . .	626
14.2.1.9	U–V Decoder Scenario . . . . .	627
14.2.1.10	U–U Decoder Scenario . . . . .	627
14.2.1.11	V–U Decoder Scenario . . . . .	627
14.2.1.12	V–V Decoder Scenario . . . . .	627
14.3	Zinc Function Modelling . . . . .	627
14.3.1	Error Minimisation . . . . .	628
14.3.2	Computational Complexity . . . . .	629
14.3.3	Reducing the Complexity of Zinc Function Excitation Optimisation . . . . .	630
14.3.4	Phases of the Zinc Functions . . . . .	631
14.4	Pitch Detection . . . . .	631
14.4.1	Voiced–unvoiced Boundaries . . . . .	632
14.4.2	Pitch Prototype Selection . . . . .	633
14.5	Voiced Speech . . . . .	635
14.5.1	Energy Scaling . . . . .	636
14.5.2	Quantisation . . . . .	638
14.6	Excitation Interpolation Between Prototype Segments . . . . .	639
14.6.1	ZFE Interpolation Regions . . . . .	640
14.6.2	ZFE Amplitude Parameter Interpolation . . . . .	642
14.6.3	ZFE Position Parameter Interpolation . . . . .	642
14.6.4	Implicit Signalling of Prototype Zero Crossing . . . . .	644
14.6.5	Removal of ZFE Pulse Position Signalling and Interpolation . . . . .	644

14.6.6	Pitch Synchronous Interpolation of Line Spectrum Frequencies . . .	645
14.6.7	ZFE Interpolation Example . . . . .	645
14.7	Unvoiced Speech . . . . .	645
14.8	Adaptive Postfilter . . . . .	645
14.9	Results for Single Zinc Function Excitation . . . . .	646
14.10	Error Sensitivity of the 1.9 kbps PWI-ZFE Coder . . . . .	649
14.10.1	Parameter Sensitivity of the 1.9 kbps PWI-ZFE Coder . . . . .	650
14.10.1.1	Line Spectrum Frequencies . . . . .	650
14.10.1.2	Voiced–unvoiced Flag . . . . .	650
14.10.1.3	Pitch Period . . . . .	651
14.10.1.4	Excitation Amplitude Parameters . . . . .	651
14.10.1.5	Root Mean Square Energy Parameter . . . . .	651
14.10.1.6	Boundary Shift Parameter . . . . .	651
14.10.2	Degradation from Bit Corruption . . . . .	652
14.10.2.1	Error Sensitivity Classes . . . . .	653
14.11	Multiple Zinc Function Excitation . . . . .	654
14.11.1	Encoding Algorithm . . . . .	654
14.11.2	Performance of Multiple Zinc Function Excitation . . . . .	657
14.12	A Sixth-rate, 3.8 kbps GSM-like Speech Transceiver . . . . .	661
14.12.1	Motivation . . . . .	661
14.12.2	The Turbo-coded Sixth-rate 3.8 kbps GSM-like System . . . . .	662
14.12.3	Turbo Channel Coding . . . . .	662
14.12.4	The Turbo-coded GMSK Transceiver . . . . .	664
14.12.5	System Performance Results . . . . .	665
14.13	Chapter Summary . . . . .	665
<b>15</b>	<b>Mixed-multiband Excitation</b>	<b>667</b>
15.1	Introduction . . . . .	667
15.2	Overview of Mixed-multiband Excitation . . . . .	668
15.3	Finite Impulse Response Filter . . . . .	671
15.4	Mixed-multiband Excitation Encoder . . . . .	673
15.4.1	Voicing Strengths . . . . .	674
15.5	Mixed-multiband Excitation Decoder . . . . .	676
15.5.1	Adaptive Postfilter . . . . .	678
15.5.2	Computational Complexity . . . . .	679
15.6	Performance of the Mixed-multiband Excitation Coder . . . . .	680
15.6.1	Performance of a Mixed-multiband Excitation Linear Predictive Coder . . . . .	680
15.6.2	Performance of a Mixed-multiband Excitation and Zinc Function Prototype Excitation Coder . . . . .	683
15.7	A Higher Rate 3.85 kbps Mixed-multiband Excitation Scheme . . . . .	686
15.8	A 2.35 kbps Joint-detection-based CDMA Speech Transceiver . . . . .	691
15.8.1	Background . . . . .	691
15.8.2	The Speech Codec’s Bit Allocation . . . . .	692
15.8.3	The Speech Codec’s Error Sensitivity . . . . .	693
15.8.4	Channel Coding . . . . .	694

15.8.5	The JD-CDMA Speech System . . . . .	695
15.8.6	System Performance . . . . .	696
15.8.7	Conclusions on the JD-CDMA Speech Transceiver . . . . .	699
15.9	Chapter Summary . . . . .	699
<b>16</b>	<b>Sinusoidal Transform Coding Below 4 kbps</b>	<b>701</b>
16.1	Introduction . . . . .	701
16.2	Sinusoidal Analysis of Speech Signals . . . . .	702
16.2.1	Sinusoidal Analysis with Peak-picking . . . . .	702
16.2.2	Sinusoidal Analysis using Analysis-by-synthesis . . . . .	703
16.3	Sinusoidal Synthesis of Speech Signals . . . . .	704
16.3.1	Frequency, Amplitude and Phase Interpolation . . . . .	704
16.3.2	Overlap-add Interpolation . . . . .	705
16.4	Low-bitrate Sinusoidal Coders . . . . .	705
16.4.1	Increased Frame Length . . . . .	708
16.4.2	Incorporating Linear Prediction Analysis . . . . .	708
16.5	Incorporating Prototype Waveform Interpolation . . . . .	709
16.6	Encoding the Sinusoidal Frequency Component . . . . .	710
16.7	Determining the Excitation Components . . . . .	712
16.7.1	Peak-picking of the Residual Spectra . . . . .	712
16.7.2	Analysis-by-synthesis of the Residual Spectrum . . . . .	713
16.7.3	Computational Complexity . . . . .	715
16.7.4	Reducing the Computational Complexity . . . . .	715
16.8	Quantising the Excitation Parameters . . . . .	720
16.8.1	Encoding the Sinusoidal Amplitudes . . . . .	720
16.8.1.1	Vector Quantisation of the Amplitudes . . . . .	720
16.8.1.2	Interpolation and Decimation . . . . .	720
16.8.1.3	Vector Quantisation . . . . .	721
16.8.1.4	Vector Quantisation Performance . . . . .	723
16.8.1.5	Scalar Quantisation of the Amplitudes . . . . .	724
16.8.2	Encoding the Sinusoidal Phases . . . . .	725
16.8.2.1	Vector Quantisation of the Phases . . . . .	725
16.8.2.2	Encoding the Phases with a Voiced–unvoiced Switch . . . . .	725
16.8.3	Encoding the Sinusoidal Fourier Coefficients . . . . .	726
16.8.3.1	Equivalent Rectangular Bandwidth Scale . . . . .	726
16.8.4	Voiced–unvoiced Flag . . . . .	727
16.9	Sinusoidal Transform Decoder . . . . .	728
16.9.1	Pitch Synchronous Interpolation . . . . .	729
16.9.1.1	Fourier Coefficient Interpolation . . . . .	729
16.9.2	Frequency Interpolation . . . . .	729
16.9.3	Computational Complexity . . . . .	729
16.10	Speech Coder Performance . . . . .	730
16.11	Chapter Summary . . . . .	736

---

<b>17</b>	<b>Conclusions on Low-rate Coding</b>	<b>737</b>
17.1	Summary . . . . .	737
17.2	Listening Tests . . . . .	738
17.3	Summary of Very-low-rate Coding . . . . .	739
17.4	Further Research . . . . .	741
<b>18</b>	<b>Comparison of Speech Codecs and Transceivers</b>	<b>743</b>
18.1	Background to Speech Quality Evaluation . . . . .	743
18.2	Objective Speech Quality Measures . . . . .	744
18.2.1	Introduction . . . . .	744
18.2.2	Signal-to-noise Ratios . . . . .	745
18.2.3	Articulation Index . . . . .	745
18.2.4	Cepstral Distance . . . . .	746
18.2.5	Example: Computation of Cepstral Coefficients . . . . .	750
18.2.6	Logarithmic Likelihood Ratio . . . . .	751
18.2.7	Euclidean Distance . . . . .	752
18.3	Subjective Measures . . . . .	752
18.3.1	Quality Tests . . . . .	752
18.4	Comparison of Subjective and Objective Measures . . . . .	753
18.4.1	Background . . . . .	753
18.4.2	Intelligibility Tests . . . . .	755
18.5	Subjective Speech Quality of Various Codecs . . . . .	755
18.6	Error Sensitivity Comparison of Various Codecs . . . . .	757
18.7	Objective Speech Performance of Various Transceivers . . . . .	757
18.8	Chapter Summary . . . . .	764
<b>19</b>	<b>The Voice over Internet Protocol</b>	<b>765</b>
19.1	Introduction . . . . .	765
19.2	Session Initiation Protocol . . . . .	766
19.2.1	Introduction . . . . .	766
19.2.2	SIP Signalling . . . . .	766
19.2.2.1	Registration . . . . .	766
19.2.2.2	Call Setup . . . . .	768
19.2.2.3	Terminate a Call . . . . .	770
19.2.2.4	Cancel a Call . . . . .	771
19.2.3	Session Description Protocol . . . . .	772
19.3	H.323 Standards . . . . .	774
19.3.1	Introduction . . . . .	774
19.3.2	H.323 Signalling . . . . .	775
19.3.2.1	Registration . . . . .	775
19.3.2.2	Call Establishment . . . . .	775
19.3.2.3	Capability Exchange . . . . .	777
19.3.2.4	Establishment of Media Communication . . . . .	777
19.3.2.5	Call Termination . . . . .	777
19.4	Real-time Transport Protocol . . . . .	778
19.4.1	RTP Header Format . . . . .	779

---

19.4.2 RTP Profiles and Payloads . . . . .	779
19.4.2.1 RTP Payload for G.711 . . . . .	779
19.4.2.2 RTP Payload for G.729 . . . . .	779
19.5 Conclusion . . . . .	781
<b>A Constructing the Quadratic Spline Wavelets</b>	<b>783</b>
<b>B Zinc Function Excitation</b>	<b>787</b>
<b>C Probability Density Function for Amplitudes</b>	<b>793</b>
<b>Bibliography</b>	<b>797</b>
<b>Index</b>	<b>825</b>
<b>Author Index</b>	<b>834</b>