Voice and Audio Compression for Wireless Communications

by

©L. Hanzo, F.C.A. Somerville, J.P. Woodard, H-T. How
School of Electronics and Computer Science,
University of Southampton, UK

Contents

Preface and Motivation **1**

Acknowledgements **11**

I Speech Signals and Waveform Coding **13**

1 Speech Signals and Coding **15**
1.1 Motivation of Speech Compression 15
1.2 Basic Characterisation of Speech Signals 16
1.3 Classification of Speech Codecs 20
1.3.1 Waveform Coding 20
1.3.1.1 Time-domain Waveform Coding 21
1.3.1.2 Frequency-domain Waveform Coding 21
1.3.2 Vocoders 22
1.3.3 Hybrid Coding 23
1.4 Waveform Coding 23
1.4.1 Digitisation of Speech 23
1.4.2 Quantisation Characteristics 25
1.4.3 Quantisation Noise and Rate-Distortion Theory 25
1.4.4 Non-uniform Quantisation for a Known PDF: Companding 28
1.4.5 PDF-independent Quantisation using Logarithmic Compression . . . 31
1.4.5.1 The μ-Law Compander 32
1.4.5.2 The A-law Compander 33
1.4.6 Optimum Non-uniform Quantisation 35
1.5 Chapter Summary 39

2 Predictive Coding **41**
2.1 Forward Predictive Coding 41
2.2 DPCM Codec Schematic 42
2.3 Predictor Design 43

2.3.1 Problem Formulation 43
2.3.2 Covariance Coefficient Computation 45
2.3.3 Predictor Coefficient Computation 46
2.4 Adaptive One-word-memory Quantization 50
2.5 DPCM Performance 53
2.6 Backward-Adaptive Prediction 55
2.6.1 Background 55
2.6.2 Stochastic Model Processes 57
2.7 The 32 kbps G.721 ADPCM Codec 60
2.7.1 Functional Description of the G.721 Codec 60
2.7.2 Adaptive Quantiser 62
2.7.3 G.721 Quantiser Scale Factor Adaptation 62
2.7.4 G.721 Adaptation Speed Control 63
2.7.5 G.721 Adaptive Prediction and Signal Reconstruction 64
2.8 Speech Quality Evaluation 66
2.9 G.726 and G.727 ADPCM Coding 68
2.9.1 Motivation 68
2.9.2 Embedded G.727 ADPCM coding 68
2.9.3 Performance of the Embedded G.727 ADPCM Codec 70
2.10 Rate-Distortion in Predictive Coding 74
2.11 Chapter Summary 80

II Analysis by Synthesis Coding 83

3 Analysis-by-synthesis Principles 85
3.1 Motivation 85
3.2 Analysis-by-synthesis Codec Structure 86
3.3 The Short-term Synthesis Filter 87
3.4 Long-Term Prediction 90
3.4.1 Open-loop Optimisation of LTP parameters 90
3.4.2 Closed-loop Optimisation of LTP parameters 96
3.5 Excitation Models 100
3.6 Adaptive Post-filtering 102
3.7 Lattice-based Linear Prediction 105
3.8 Chapter Summary 111

4 Speech Spectral Quantization 113
4.1 Log-area Ratios 113
4.2 Line Spectral Frequencies 117
4.2.1 Derivation of the Line Spectral Frequencies 117
4.2.2 Computation of the Line Spectral Frequencies 121
4.2.3 Chebyshev-description of Line Spectral Frequencies 123
4.3 Spectral Vector Quantization 125
4.3.1 Background 125
4.3.2 Speaker-adaptive Vector Quantisation of LSFs 129

4.3.3 Stochastic VQ of LPC Parameters . . . 130
4.3.3.1 Background . . . 131
4.3.3.2 The Stochastic VQ Algorithm . . . 132
4.3.4 Robust Vector Quantisation Schemes for LSFs . . . 134
4.3.5 LSF Vector-quantisers in Standard Codecs . . . 136
4.4 Spectral Quantizers for Wideband Speech Coding . . . 137
4.4.1 Introduction to Wideband Spectral Quantisation . . . 137
4.4.1.1 Statistical Properties of Wideband LSFs . . . 139
4.4.1.2 Speech Codec Specifications . . . 139
4.4.2 Wideband LSF Vector Quantizers . . . 142
4.4.2.1 Memoryless Vector Quantization . . . 142
4.4.2.2 Predictive Vector Quantization . . . 145
4.4.2.3 Multimode Vector Quantization . . . 149
4.4.3 Simulation Results and Subjective Evaluations . . . 152
4.4.4 Conclusions on Wideband Spectral Quantisation . . . 153
4.5 Chapter Summary . . . 154

5 RPE Coding 155
5.1 Theoretical Background . . . 155
5.2 The 13 kbps RPE-LTP GSM Speech encoder . . . 162
5.2.1 Pre-processing . . . 162
5.2.2 STP analysis filtering . . . 164
5.2.3 LTP analysis filtering . . . 165
5.2.4 Regular Excitation Pulse Computation . . . 165
5.3 The 13 kbps RPE-LTP GSM Speech Decoder . . . 166
5.4 Bit-sensitivity of the GSM Codec . . . 170
5.5 A 'Tool-box' Based Speech Transceiver . . . 171
5.6 Chapter Summary . . . 172

6 Forward-Adaptive CELP Coding 175
6.1 Background . . . 175
6.2 The Original CELP Approach . . . 176
6.3 Fixed Codebook Search . . . 179
6.4 CELP Excitation Models . . . 181
6.4.1 Binary Pulse Excitation . . . 181
6.4.2 Transformed Binary Pulse Excitation . . . 182
6.4.2.1 Excitation Generation . . . 182
6.4.2.2 TBPE Bit Sensitivity . . . 184
6.4.3 Dual-rate Algebraic CELP Coding . . . 187
6.4.3.1 ACELP Codebook Structure . . . 187
6.4.3.2 Dual-rate ACELP Bitallocation . . . 189
6.4.3.3 Dual-rate ACELP Codec Performance . . . 190
6.5 CELP Optimization . . . 191
6.5.1 Introduction . . . 191
6.5.2 Calculation of the Excitation Parameters . . . 192
6.5.2.1 Full Codebook Search Theory . . . 192

6.5.2.2 Sequential Search Procedure . . . 194
6.5.2.3 Full Search Procedure . . . 195
6.5.2.4 Sub-Optimal Search Procedures . . . 197
6.5.2.5 Quantization of the Codebook Gains . . . 198
6.5.3 Calculation of the Synthesis Filter Parameters . . . 200
6.5.3.1 Bandwidth Expansion . . . 201
6.5.3.2 Least Squares Techniques . . . 201
6.5.3.3 Optimization via Powell's Method . . . 204
6.5.3.4 Simulated Annealing and the Effects of Quantization . . . 205
6.6 CELP Error-sensitivity . . . 209
6.6.1 Introduction . . . 209
6.6.2 Improving the Spectral Information Error Sensitivity . . . 209
6.6.2.1 LSF Ordering Policies . . . 209
6.6.2.2 The Effect of FEC on the Spectral Parameters . . . 211
6.6.2.3 The Effect of Interpolation . . . 212
6.6.3 Improving the Error Sensitivity of the Excitation Parameters . . . 213
6.6.3.1 The Fixed Codebook Index . . . 214
6.6.3.2 The Fixed Codebook Gain . . . 214
6.6.3.3 Adaptive Codebook Delay . . . 215
6.6.3.4 Adaptive Codebook Gain . . . 215
6.6.4 Matching Channel Codecs to the Speech Codec . . . 216
6.6.5 Error Resilience Conclusions . . . 220
6.7 Dual-mode Speech Transceiver . . . 221
6.7.1 The Transceiver Scheme . . . 221
6.7.2 Re-configurable Modulation . . . 222
6.7.3 Source-matched Error Protection . . . 224
6.7.3.1 Low-quality 3.1 kBd Mode . . . 224
6.7.3.2 High-quality 3.1 kBd Mode . . . 228
6.7.4 Packet Reservation Multiple Access . . . 229
6.7.5 3.1 kBd System Performance . . . 231
6.7.6 3.1 kBd System Summary . . . 234
6.8 Multi-slot PRMA Transceiver . . . 235
6.8.1 Background and Motivation . . . 235
6.8.2 PRMA-assisted Multi-slot Adaptive Modulation . . . 235
6.8.3 Adaptive GSM-like Schemes . . . 237
6.8.4 Adaptive DECT-like Schemes . . . 238
6.8.5 Summary of Adaptive Multi-slot PRMA . . . 239
6.9 Chapter Summary . . . 240

7 Standard Speech Codecs 241
7.1 Background . . . 241
7.2 The US DoD FS-1016 4.8 kbits/s CELP Codec . . . 241
7.2.1 Introduction . . . 241
7.2.2 LPC Analysis and Quantization . . . 243
7.2.3 The Adaptive Codebook . . . 244
7.2.4 The Fixed Codebook . . . 245

7.2.5 Error Concealment Techniques 246
7.2.6 Decoder Post-Filtering 247
7.2.7 Conclusion 247
7.3 The IS-54 DAMPS speech codec 247
7.4 The JDC speech codec 251
7.5 The Qualcomm Variable Rate CELP Codec 253
7.5.1 Introduction 253
7.5.2 Codec Schematic and Bit Allocation 254
7.5.3 Codec Rate Selection 255
7.5.4 LPC Analysis and Quantization 256
7.5.5 The Pitch Filter 257
7.5.6 The Fixed Codebook 258
7.5.7 Rate 1/8 Filter Excitation 259
7.5.8 Decoder Post-Filtering 260
7.5.9 Error Protection and Concealment Techniques 260
7.5.10 Conclusion 261
7.6 Japanese Half-Rate Speech Codec 261
7.6.1 Introduction 261
7.6.2 Codec Schematic and Bit Allocation 262
7.6.3 Encoder Pre-Processing 264
7.6.4 LPC Analysis and Quantization 264
7.6.5 The Weighting Filter 265
7.6.6 Excitation Vector 1 265
7.6.7 Excitation Vector 2 266
7.6.8 Channel Coding 266
7.6.9 Decoder Post Processing 268
7.7 The half-rate GSM codec 269
7.7.1 Half-rate GSM codec outline 269
7.7.2 Half-rate GSM Codec's Spectral Quantisation 271
7.7.3 Error protection 272
7.8 The 8 kbits/s G.729 Codec 273
7.8.1 Introduction 273
7.8.2 Codec Schematic and Bit Allocation 274
7.8.3 Encoder Pre-Processing 275
7.8.4 LPC Analysis and Quantization 276
7.8.5 The Weighting Filter 278
7.8.6 The Adaptive Codebook 279
7.8.7 The Fixed Algebraic Codebook 280
7.8.8 Quantization of the Gains 283
7.8.9 Decoder Post Processing 284
7.8.10 G.729 Error Concealment Techniques 286
7.8.11 G.729 Bit-sensitivity 287
7.8.12 Turbo-coded OFDM G.729 Speech Transceiver 288
7.8.12.1 Background 288
7.8.12.2 System Overview 288
7.8.12.3 Turbo Channel Encoding 289

7.8.12.4 OFDM in the FRAMES Speech/Data Sub–Burst 290
7.8.12.5 Channel model 290
7.8.12.6 Turbo-coded G.729 OFDM Parameters 291
7.8.12.7 Turbo-coded G.729 OFDM Performance 292
7.8.12.8 Turbo-coded G.729 OFDM Summary 293
7.8.13 G.729 Summary 295
7.9 The Reduced Complexity G.729 Annex A Codec 295
7.9.1 Introduction 295
7.9.2 The Perceptual Weighting Filter 296
7.9.3 The Open Loop Pitch Search 296
7.9.4 The Closed Loop Pitch Search 296
7.9.5 The Algebraic Codebook Search 297
7.9.6 The Decoder Post Processing 298
7.9.7 Conclusions 298
7.10 The Enhanced Full-rate GSM codec 298
7.10.1 Codec Outline 298
7.10.2 Operation of the EFR-GSM Encoder 300
7.10.2.1 Spectral Quantisation in the EFR-GSM Codec 300
7.10.2.2 Adaptive Codebook Search 302
7.10.2.3 Fixed Codebook Search 303
7.11 The IS-136 Speech Codec 304
7.11.1 IS-136 codec outline 304
7.11.2 IS-136 Bitallocation scheme 305
7.11.3 Fixed Codebook Search 307
7.11.4 IS-136 Channel Coding 307
7.12 The ITU G.723.1 Dual-Rate Codec 308
7.12.1 Introduction 308
7.12.2 G.723.1 Encoding Principle 309
7.12.3 Vector-Quantisation of the LSPs 312
7.12.4 Formant-based Weighting Filter 312
7.12.5 The 6.3 kbps High-rate G.723.1 Excitation 313
7.12.6 The 5.3 kbps low-rate G.723.1 excitation 314
7.12.7 G.723.1 Bitallocation 315
7.12.8 G.723.1 Error Sensitivity 317
7.13 Advanced Multi-rate JD-CDMA Transceiver 319
7.13.1 Multi-rate codecs and systems 319
7.13.2 System Overview 322
7.13.3 The Adaptive Multi-Rate Speech Codec 323
7.13.3.1 AMR Codec Overview 323
7.13.3.2 Linear Prediction Analysis 324
7.13.3.3 LSF Quantization 324
7.13.3.4 Pitch Analysis 325
7.13.3.5 Fixed Codebook With Algebraic Structure 326
7.13.3.6 Post-Processing 327
7.13.3.7 The AMR Codec's Bit Allocation 327
7.13.3.8 Codec Mode Switching Philosophy 328

7.13.4 The AMR Speech Codec's Error Sensitivity 328
7.13.5 Redundant Residue Number System Based Channel Coding 332
7.13.5.1 Redundant Residue Number System Overview 332
7.13.5.2 Source-Matched Error Protection 333
7.13.6 Joint Detection Code Division Multiple Access 335
7.13.6.1 Overview 335
7.13.6.2 Joint Detection Based Adaptive Code Division Multiple Access 336
7.13.7 System Performance 336
7.13.7.1 Subjective Testing 342
7.13.8 Conclusions 345
7.14 Chapter Summary 345

8 Backward-Adaptive CELP Coding 349
8.1 Introduction 349
8.2 Motivation and Background 350
8.3 Backward-Adaptive G728 Schematic 352
8.4 Backward-Adaptive G728 Coding 355
8.4.1 G728 Error Weighting 355
8.4.2 G728 Windowing 355
8.4.3 Codebook Gain Adaption 359
8.4.4 G728 Codebook Search 361
8.4.5 G728 Excitation Vector Quantization 364
8.4.6 G728 Adaptive Postfiltering 366
8.4.6.1 Adaptive Long-term Postfiltering 366
8.4.6.2 G728 Adaptive Short-term Postfiltering 369
8.4.7 Complexity and Performance of the G728 Codec 369
8.5 Reduced-Rate 16-8 kbps G728-Like Codec I 370
8.6 The Effects of Long Term Prediction 373
8.7 Closed-Loop Codebook Training 378
8.8 Reduced-Rate 16-8 kbps G728-Like Codec II 383
8.9 Programmable-Rate 8-4 kbps CELP Codecs 383
8.9.1 Motivation 383
8.9.2 8-4kbps Codec Improvements 384
8.9.3 8-4kbps Codecs - Forward Adaption of the STP Synthesis Filter . . . 385
8.9.4 8-4kbps Codecs - Forward Adaption of the LTP 387
8.9.4.1 Initial Experiments 387
8.9.4.2 Quantization of Jointly Optimized Gains 389
8.9.4.3 8-4kbps Codecs - Voiced/Unvoiced Codebooks 392
8.9.5 Low Delay Codecs at 4-8 kbits/s 393
8.9.6 Low Delay ACELP Codec 397
8.10 Backward-adaptive Error Sensitivity Issues 400
8.10.1 The Error Sensitivity of the G728 Codec 400
8.10.2 The Error Sensitivity of Our 4-8 kbits/s Low Delay Codecs 401
8.10.3 The Error Sensitivity of Our Low Delay ACELP Codec 406
8.11 A Low-Delay Multimode Speech Transceiver 407

8.11.1 Background 407
8.11.2 8-16 kbps Codec Performance 408
8.11.3 Transmission Issues 410
8.11.3.1 Higher-quality Mode 410
8.11.3.2 Lower-quality Mode 411
8.11.4 Speech Transceiver Performance 411
8.12 Chapter Summary 412

III Wideband Coding and Transmission 413

9 Wideband Speech Coding 415
9.1 Subband-ADPCM Wideband Coding 415
9.1.1 Introduction and Specifications 415
9.1.2 G722 Codec Outline 416
9.1.3 Principles of Subband Coding 419
9.1.4 Quadrature Mirror Filtering 421
9.1.4.1 Analysis Filtering 421
9.1.4.2 Synthesis Filtering 423
9.1.4.3 Practical QMF Design Constraints 425
9.1.5 G722 Adaptive Quantisation and Prediction 431
9.1.6 G722 Coding Performance 433
9.2 Wideband Transform-Coding at 32 kbps 433
9.2.1 Background 433
9.2.2 Transform-Coding Algorithm 433
9.3 Subband-Split Wideband CELP Codecs 437
9.3.1 Background 437
9.3.2 Subband-based Wideband CELP coding 437
9.3.2.1 Motivation 437
9.3.2.2 Low-band Coding 439
9.3.2.3 Highband Coding 439
9.3.2.4 Bit allocation Scheme 439
9.4 Fullband Wideband ACELP Coding 440
9.4.1 Wideband ACELP Excitation 440
9.4.2 Wideband 32 kbps ACELP Coding 443
9.4.3 Wideband 9.6 kbps ACELP Coding 444
9.5 Turbo-coded Wideband Speech Transceiver 445
9.5.1 Background and Motivation 445
9.5.2 System Overview 448
9.5.3 System Parameters 449
9.5.4 Constant Throughput Adaptive Modulation 450
9.5.5 Adaptive Wideband Transceiver Performance 451
9.5.6 Multi–mode Transceiver Adaptation 452
9.5.7 Transceiver Mode Switching 454
9.5.8 The Wideband G.722.1 Codec 456
9.5.8.1 Audio Codec Overview 456

9.5.9 Detailed Description of the Audio Codec . . . 457
9.5.10 Wideband Adaptive System Performance . . . 459
9.5.11 Audio Frame Error Results . . . 459
9.5.12 Audio Segmental SNR Performance and Discussions . . . 461
9.5.13 G.722.1 Audio Transceiver Summary and Conclusions . . . 462
9.6 Turbo-Detected IRCC AMR-WB Transceivers . . . 463
9.6.1 Introduction . . . 463
9.6.2 The AMR-WB Codec's Error Sensitivity . . . 465
9.6.3 System Model . . . 466
9.6.4 Design of Irregular Convolutional Codes . . . 467
9.6.5 An Example Irregular Convolutional Code . . . 469
9.6.6 UEP AMR IRCC Performance Results . . . 470
9.6.7 UEP AMR Conclusions . . . 472
9.7 The AMR-WB+ Audio Codec . . . 474
9.7.1 Introduction . . . 474
9.7.2 Audio requirements in mobile multimedia applications . . . 477
9.7.2.1 Summary of audio-visual services . . . 478
9.7.2.2 Bit rates supported by the radio network . . . 478
9.7.3 Overview of the AMR-WB+ codec . . . 478
9.7.3.1 Encoding the high frequencies . . . 482
9.7.3.2 Stereo encoding . . . 482
9.7.3.3 Complexity of AMR-WB+ . . . 483
9.7.3.4 Transport and file format of AMR-WB+ . . . 483
9.7.4 Performance of AMR-WB+ . . . 484
9.7.5 Summary of the AMR-WB+ codec . . . 486
9.8 Chapter Summary . . . 487

10 Advanced Multi-Rate Speech Transceivers 489
10.1 Introduction . . . 489
10.2 The Adaptive Multi-Rate Speech Codec . . . 491
10.2.1 Overview . . . 491
10.2.2 Linear Prediction Analysis . . . 493
10.2.3 LSF Quantization . . . 493
10.2.4 Pitch Analysis . . . 493
10.2.5 Fixed Codebook With Algebraic Structure . . . 495
10.2.6 Post-Processing . . . 496
10.2.7 The AMR Codec's Bit Allocation . . . 496
10.2.8 Codec Mode Switching Philosophy . . . 496
10.3 Speech Codec's Error Sensitivity . . . 497
10.4 System Background . . . 501
10.5 System Overview . . . 503
10.6 Redundant Residue Number System (RRNS) Channel Coding . . . 504
10.6.1 Overview . . . 504
10.6.2 Source-Matched Error Protection . . . 505
10.7 Joint Detection Code Division Multiple Access . . . 508
10.7.1 Overview . . . 508

10.7.2 Joint Detection Based Adaptive Code Division Multiple Access . . . 508
10.8 System Performance . . . 509
10.8.1 Subjective Testing . . . 518
10.9 A Turbo-Detected Irregular Convolutional Coded AMR Transceiver . . . 519
10.9.1 Motivation . . . 519
10.9.2 The AMR-WB Codec's Error Sensitivity . . . 520
10.9.3 System Model . . . 520
10.9.4 Design of Irregular Convolutional Codes . . . 522
10.9.5 An Example Irregular Convolutional Code . . . 524
10.9.6 UEP AMR IRCC Performance Results . . . 525
10.9.7 UEP AMR Conclusions . . . 527
10.10 Chapter Summary . . . 529

11 MPEG-4 Audio Compression and Transmission 531
11.1 Overview of MPEG-4 Audio . . . 531
11.2 General Audio Coding . . . 533
11.2.1 Advanced Audio Coding . . . 541
11.2.2 Gain Control Tool . . . 544
11.2.3 Psychoacoustic Model . . . 545
11.2.4 Temporal Noise Shaping . . . 547
11.2.5 Stereophonic Coding . . . 549
11.2.6 AAC Quantization and Coding . . . 550
11.2.7 Noiseless Huffman Coding . . . 552
11.2.8 Bit-Sliced Arithmetic Coding . . . 553
11.2.9 Transform-domain Weighted Interleaved Vector Quantization . . . 555
11.2.10 Parametric Audio Coding . . . 558
11.3 Speech Coding in MPEG-4 Audio . . . 559
11.3.1 Harmonic Vector Excitation Coding . . . 559
11.3.2 CELP Coding in MPEG-4 . . . 562
11.3.3 LPC Analysis and Quantization . . . 564
11.3.4 Multi Pulse and Regular Pulse Excitation . . . 565
11.4 MPEG-4 Codec Performance . . . 567
11.5 MPEG-4 Space-Time Block Coded OFDM Audio Transceiver . . . 569
11.5.1 System Overview . . . 571
11.5.2 System parameters . . . 571
11.5.3 Frame Dropping Procedure . . . 572
11.5.4 Space-Time Coding . . . 575
11.5.5 Adaptive Modulation . . . 576
11.5.6 System Performance . . . 579
11.6 Turbo-Detected STTC Aided MPEG-4 Audio Transceivers . . . 581
11.6.1 Motivation and Background . . . 581
11.6.2 Audio Turbo Transceiver Overview . . . 583
11.6.3 The Turbo Transceiver . . . 584
11.6.4 Turbo Transceiver Performance Results . . . 586
11.6.5 MPEG-4 Turbo Transceiver Summary . . . 589
11.7 Turbo-Detected STTC Aided MPEG-4 Versus AMR-WB Transceivers . . . 590

11.7.1 Motivation and Background 590
11.7.2 The AMR-WB Codec'S Error Sensitivity 591
11.7.3 The MPEG-4 TwinVQ Codec'S Error Sensitivity 593
11.7.4 The Turbo Transceiver 594
11.7.5 Performance Results 596
11.7.6 AMR-WB and MPEG-4 TwinVQ Turbo Transceiver Summary 599
11.8 Chapter Summary 599

IV Very Low Rate Coding and Transmission 601

12 Overview of Low-rate Speech Coding 603
12.1 Low Bit Rate Speech Coding 603
12.1.1 Analysis-by-Synthesis Coding 605
12.1.2 Speech Coding at 2.4kbps 607
12.1.2.1 Background to 2.4kbps Speech Coding 608
12.1.2.2 Frequency Selective Harmonic Coder 609
12.1.2.3 Sinusoidal Transform Coder 610
12.1.2.4 Multiband Excitation Coders 611
12.1.2.5 Subband Linear Prediction Coder 612
12.1.2.6 Mixed Excitation Linear Prediction Coder 613
12.1.2.7 Waveform Interpolation Coder 615
12.1.3 Speech Coding Below 2.4kbps 616
12.2 Linear Predictive Coding model 617
12.2.1 Short Term Prediction 618
12.2.2 Long Term Prediction 619
12.2.3 Final Analysis-by-Synthesis Model 620
12.3 Speech Quality Measurements 620
12.3.1 Objective Speech Quality Measures 621
12.3.2 Subjective Speech Quality Measures 622
12.3.3 2.4kbps Selection Process 622
12.4 Speech Database 624
12.5 Chapter Summary 625

13 Linear Predictive Vocoder 629
13.1 Overview of a Linear Predictive Vocoder 629
13.2 Line Spectrum Frequencies Quantization 630
13.2.1 Line Spectrum Frequencies Scalar Quantization 630
13.2.2 Line Spectrum Frequencies Vector Quantization 631
13.3 Pitch Detection 635
13.3.1 Voiced-Unvoiced Decision 637
13.3.2 Oversampled Pitch Detector 638
13.3.3 Pitch Tracking 641
13.3.3.1 Computational Complexity 644
13.3.4 Integer Pitch Detector 646
13.4 Unvoiced Frames 647

13.5 Voiced Frames 648
13.5.1 Placement of Excitation Pulses 648
13.5.2 Pulse Energy 649
13.6 Adaptive Postfilter 649
13.7 Pulse Dispersion Filter 652
13.7.1 Pulse Dispersion Principles 652
13.7.2 Pitch Independent Glottal Pulse Shaping Filter 653
13.7.3 Pitch Dependent Glottal Pulse Shaping Filter 654
13.8 Results for Linear Predictive Vocoder 655
13.9 Chapter Summary 660

14 Wavelets and Pitch Detection 661
14.1 Conceptual Introduction to Wavelets 661
14.1.1 Fourier Theory 661
14.1.2 Wavelet Theory 662
14.1.3 Detecting Discontinuities with Wavelets 663
14.2 Introduction to Wavelet Mathematics 664
14.2.1 Multiresolution Analysis 665
14.2.2 Polynomial Spline Wavelets 666
14.2.3 Pyramidal Algorithm 668
14.2.4 Boundary Effects 668
14.3 Preprocessing the Wavelet Transform Signal 669
14.3.1 Spurious Pulses 669
14.3.2 Normalization 672
14.3.3 Candidate Glottal Pulses 672
14.4 Voiced-Unvoiced Decision 673
14.5 Wavelet Based Pitch Detector 673
14.5.1 Dynamic Programming 674
14.5.2 Autocorrelation Simplification 677
14.6 Chapter Summary 681

15 Zinc Function Excitation 683
15.1 Introduction 683
15.2 Overview of Prototype Waveform Interpolation Zinc Function Excitation . . 684
15.2.1 Coding Scenarios 684
15.2.1.1 U-U-U Encoder Scenario 685
15.2.1.2 U-U-V Encoder Scenario 685
15.2.1.3 V-U-U Encoder Scenario 687
15.2.1.4 U-V-U Encoder Scenario 687
15.2.1.5 V-V-V Encoder Scenario 687
15.2.1.6 V-U-V Encoder Scenario 687
15.2.1.7 U-V-V Encoder Scenario 688
15.2.1.8 V-V-U Encoder Scenario 688
15.2.1.9 U-V Decoder Scenario 688
15.2.1.10 U-U Decoder Scenario 689
15.2.1.11 V-U Decoder Scenario 689

15.2.1.12 V-V Decoder Scenario 689
15.3 Zinc Function Modelling 689
15.3.1 Error Minimization 690
15.3.2 Computational Complexity 691
15.3.3 Reducing the Complexity of Zinc Function Excitation Optimization . 692
15.3.4 Phases of the Zinc Functions 693
15.4 Pitch Detection 693
15.4.1 Voiced-Unvoiced Boundaries 693
15.4.2 Pitch Prototype Selection 694
15.5 Voiced Speech 696
15.5.1 Energy Scaling 699
15.5.2 Quantization 699
15.6 Excitation Interpolation Between Prototype Segments 701
15.6.1 ZFE Interpolation Regions 701
15.6.2 ZFE Amplitude Parameter Interpolation 702
15.6.3 ZFE Position Parameter Interpolation 702
15.6.4 Implicit Signalling of Prototype Zero Crossing 704
15.6.5 Removal of ZFE Pulse Position Signalling and Interpolation 704
15.6.6 Pitch Synchronous Interpolation of Line Spectrum Frequencies . . . 705
15.6.7 ZFE Interpolation Example 705
15.7 Unvoiced Speech 705
15.8 Adaptive Postfilter 705
15.9 Results for Single Zinc Function Excitation 708
15.10 Error Sensitivity of the 1.9kbps PWI-ZFE Coder 711
15.10.1 Parameter Sensitivity of the 1.9kbps PWI-ZFE coder 711
15.10.1.1 Line Spectrum Frequencies 711
15.10.1.2 Voiced-Unvoiced Flag 712
15.10.1.3 Pitch Period 712
15.10.1.4 Excitation Amplitude Parameters 712
15.10.1.5 Root Mean Square Energy Parameter 712
15.10.1.6 Boundary Shift Parameter 713
15.10.2 Degradation from Bit Corruption 713
15.10.2.1 Error Sensitivity Classes 713
15.11 Multiple Zinc Function Excitation 715
15.11.1 Encoding Algorithm 715
15.11.2 Performance of Multiple Zinc Function Excitation 718
15.12 A Sixth-rate, 3.8 kbps GSM-like Speech Transceiver 722
15.12.1 Motivation 722
15.12.2 The Turbo-coded Sixth-rate 3.8 kbps GSM-like System 722
15.12.3 Turbo Channel Coding 723
15.12.4 The Turbo-coded GMSK Transceiver 724
15.12.5 System Performance Results 725
15.13 Chapter Summary 726

16 Mixed-Multiband Excitation **729**
16.1 Introduction . . . 729
16.2 Overview of Mixed-Multiband Excitation . . . 731
16.3 Finite Impulse Response Filter . . . 734
16.4 Mixed-Multiband Excitation Encoder . . . 735
16.4.1 Voicing Strengths . . . 737
16.5 Mixed-Multiband Excitation Decoder . . . 739
16.5.1 Adaptive Postfilter . . . 741
16.5.2 Computational Complexity . . . 741
16.6 Performance of the Mixed-Multiband Excitation Coder . . . 743
16.6.1 Performance of a Mixed-Multiband Excitation Linear Predictive Coder 743
16.6.2 Performance of a Mixed-Multiband Excitation and Zinc Function Prototype Excitation Coder . . . 748
16.7 A Higher Rate 3.85kbps Mixed-Multiband Excitation Scheme . . . 751
16.8 A 2.35 kbit/s Joint-Detection CDMA Speech Transceiver . . . 754
16.8.1 Background . . . 754
16.8.2 The Speech Codec's Bit Allocation . . . 754
16.8.3 The Speech Codec's Error Sensitivity . . . 754
16.8.4 Channel Coding . . . 755
16.8.5 The JD-CDMA Speech System . . . 756
16.8.6 System performance . . . 757
16.8.7 Conclusions on the JD-CDMA Speech Transceiver . . . 758
16.9 Chapter Summary . . . 758

17 Sinusoidal Transform Coding Below 4kbps **761**
17.1 Introduction . . . 761
17.2 Sinusoidal Analysis of Speech Signals . . . 762
17.2.1 Sinusoidal Analysis with Peak Picking . . . 762
17.2.2 Sinusoidal Analysis using Analysis-by-Synthesis . . . 763
17.3 Sinusoidal Synthesis of Speech Signals . . . 764
17.3.1 Frequency, Amplitude and Phase Interpolation . . . 764
17.3.2 Overlap-Add Interpolation . . . 765
17.4 Low Bit Rate Sinusoidal Coders . . . 765
17.4.1 Increased Frame Length . . . 768
17.4.2 Incorporating Linear Prediction Analysis . . . 768
17.5 Incorporating Prototype Waveform Interpolation . . . 769
17.6 Encoding the Sinusoidal Frequency Component . . . 770
17.7 Determining the Excitation Components . . . 773
17.7.1 Peak-Picking of the Residual Spectra . . . 773
17.7.2 Analysis-by-Synthesis of the Residual Spectrum . . . 773
17.7.3 Computational Complexity . . . 775
17.7.4 Reducing the Computational Complexity . . . 775
17.8 Quantizing the Excitation Parameters . . . 779
17.8.1 Encoding the Sinusoidal Amplitudes . . . 779
17.8.1.1 Vector Quantization of the Amplitudes . . . 780
17.8.1.2 Interpolation and Decimation . . . 780

17.8.1.3 Vector Quantization 781
17.8.1.4 Vector Quantization Performance 782
17.8.1.5 Scalar Quantization of the Amplitudes 783
17.8.2 Encoding the Sinusoidal Phases 785
17.8.2.1 Vector Quantization of the Phases 785
17.8.2.2 Encoding the Phases with a Voiced-Unvoiced Switch . . . 785
17.8.3 Encoding the Sinusoidal Fourier Coefficients 786
17.8.3.1 Equivalent Rectangular Bandwidth Scale 786
17.8.4 Voiced-Unvoiced Flag 787
17.9 Sinusoidal Transform Decoder 788
17.9.1 Pitch Synchronous Interpolation 788
17.9.1.1 Fourier Coefficient Interpolation 789
17.9.2 Frequency Interpolation 789
17.9.3 Computational Complexity 789
17.10 Speech Coder Performance 790
17.11 Chapter Summary 796

18 Conclusions on Low Rate Coding 797
18.1 Summary 797
18.2 Listening Tests 798
18.3 Summary of Very Low Rate Coding 799
18.4 Further Research 801

19 Comparison of Speech Transceivers 803
19.1 Background to Speech Quality Evaluation 803
19.2 Objective Speech Quality Measures 804
19.2.1 Introduction 804
19.2.2 Signal to Noise Ratios 805
19.2.3 Articulation Index 805
19.2.4 Ceptral Distance 806
19.2.5 Cepstral Example 809
19.2.6 Logarithmic likelihood ratio 811
19.2.7 Euclidean Distance 812
19.3 Subjective Measures 812
19.3.1 Quality Tests 813
19.4 Comparison of Quality Measures 814
19.4.1 Background 814
19.4.2 Intelligibility tests 815
19.5 Subjective Speech Quality of Various Codecs 816
19.6 Speech Codec Bit-sensitivity 818
19.7 Transceiver Speech Performance 818
19.8 Chapter Summary 825

A Constructing the Quadratic Spline Wavelets 827

B Zinc Function Excitation 831

C Probability Density Function for Amplitudes **837**

Bibliography **843**

Index **887**

Author Index **887**

Preface and Motivation

The Speech Coding Scene

Despite the emergence of sophisticated high-rate multimedia services, voice communications remain the predominant means of human communications, although the compressed voice signals may be delivered via the Internet. The large-scale, pervasive introduction of wireless Internet services is likely to promote the unified transmission of both voice and data signals using the Voice over Internet Protocol (VoIP) even in the third - generation (3G) wireless systems, despite wasting much of the valuable frequency resources for the transmission of packet headers. Even when the predicted surge of wireless data and Internet services becomes a reality, voice remains the most natural means of human communications, although this may be delivered via the Internet.

This book is dedicated to audio and voice compression issues, although the aspects of error resilience, coding delay, implementational complexity and bitrate are also at the centre of our discussions, characterising many different speech codecs incorported in source-sensitivity matched wireless transceivers. A unique feature of the book is that it also provides cutting-edge turbo-transceiver-aided research-oriented design examples and an a chapter on the VoIP protocol.

Here we attempt a rudimentary comparison of some of the codec schemes treated in the book in terms of their speech quality and bitrate, in order to provide a road map for the reader with reference to Cox's work [1, 2]. The formally evaluated Mean Opinion Score (MOS) values of the various codecs portrayed in the book are shown in Figure 1.

Observe in the figure that over the years a range of speech codecs have emerged, which attained the quality of the 64 kbps G.711 PCM speech codec, although at the cost of significantly increased coding delay and implementational complexity. The 8 kbps G.729 codec is the most recent addition to this range of the International Telecommunications Union's (ITU) standard schemes, which significantly outperforms all previous standard ITU codecs in robustness terms. The performance target of the 4 kbps ITU codec (ITU4) is also to maintain this impressive set of specifications. The family of codecs designed for various mobile radio systems - such as the 13 kbps Regular Pulse Excited (RPE) scheme of the Global System of Mobile communications known as GSM, the 7.95 kbps IS-54, and the IS-95 Pan-American schemes, the 6.7 kbps Japanese Digital Cellular (JDC) and 3.45 kbps half-rate JDC arrangement (JDC/2) - exhibits slightly lower MOS values than the ITU codecs. Let us now consider the subjective quality of these schemes in a little more depth.

The 2.4 kbps US Department of Defence Federal Standard codec known as FS-1015 is the only vocoder in this group and it has a rather synthetic speech quality, associated with the lowest subjective assessment in the figure. The 64 kbps G.711 PCM codec and the G.726/G.727 Adaptive Differential PCM (ADPCM) schemes are waveform codecs. They exhibit a low im-

plementational complexity associated with a modest bitrate economy. The remaining codecs belong to the so-called hybrid coding family and achieve significant bitrate economies at the cost of increased complexity and delay.

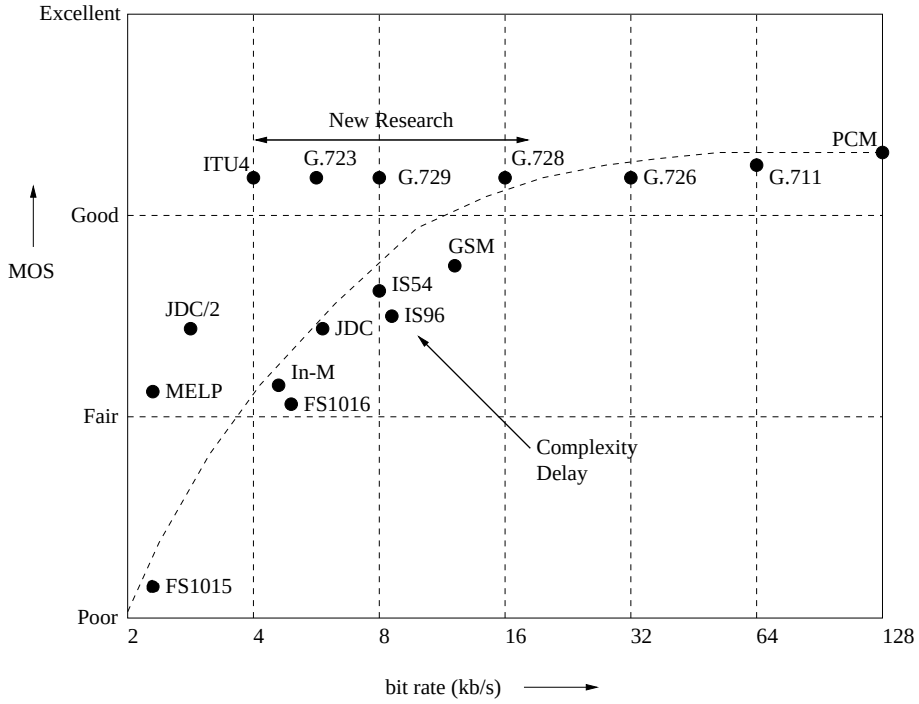

Figure 1: Subjective speech quality of various codecs [1] ©IEEE, 1996

Specifically, the 16 kbps G.728 backward-adaptive scheme maintains a similar speech quality to the 32 and 64 kbps waveform codecs, while also maintaining an impressively low, 2 ms delay. This scheme was standardised during the early nineties. The similar-quality, but significantly more robust 8 kbps G.729 codec was approved in March 1996 by the ITU. Its standardisation overlapped with the G.723.1 codec developments. The G.723.1 codec's 6.4 kbps mode maintains a speech quality similar to the G.711, G.726, G.727, G.728 and G.728 codecs, while its 5.3 kbps mode exhibits a speech quality similar to the cellular speech codecs of the late eighties. The standardisation of a 4 kbps ITU scheme, which we refer to here as ITU4 is also a desirable design goal at the time of writing.

In parallel to the ITU's standardisation activities a range of speech coding standards have been proposed for regional cellular mobile systems. The standardisation of the 13 kbps RPE-LTP full-rate GSM (GSM-FR) codec dates back to the second half of the eighties, representing the first standard hybrid codec. Its complexity is significantly lower than that of the more recent Code Excited Linear Predictive (CELP) based codecs. Observe in the figure that there is also a similar-rate Enhanced Full-Rate GSM codec (GSM-EFR), which matches the speech quality of the G.729 and G.728 schemes. The original GSM-FR codec's development was followed a little later by the release of the 7.95 kbps Vector Sum Excited Linear Predictive

(VSELP) IS-54 American cellular standard. Due to advances in the field the 7.95 kbps IS-54 codec achieved a similar subjective speech quality to the 13 kbps GSM-FR scheme. The definition of the 6.7 kbps Japanese JDC VSELP codec was almost coincident with that of the IS-54 arrangement. This codec development was also followed by a half-rate standardisation process, leading to the 3.2 kbps Pitch-Synchroneous Innovation CELP (PSI-CELP) scheme.

The IS-95 Pan-American CDMA system also has its own standardised CELP-based speech codec, which is a variable-rate scheme, supporting bitrates between 1.2 and 14.4 kbps, depending on the prevalent voice activity. The perceived speech quality of these cellular speech codecs contrived mainly during the late eighties was found subjectively similar to each other under the perfect channel conditions of Figure 1. Lastly, the 5.6 kbps half-rate GSM codec (GSM-HR) also met its specification in terms of achieving a similar speech quality to the 13 kbps original GSM-FR arrangements, although at the cost of quadruple complexity and higher latency.

Recently the advantages of intelligent multimode speech terminals (IMT), which can reconfigure themselves in a number of different bitrate, quality and robustness modes attracted substantial research attention in the community, which led to the standardisation of the High-Speed Downlink Packet Access (HSDPA) mode of the 3G wireless systems. The HSDPA-style transceivers employ both adaptive modulation and adaptive channel coding, which result in a channel-quality dependent bit-rate fluctuation, hence requiring reconfigurable multimode voice and audio codecs, such as the Advanced Multi-Rate codec referred to as the AMR scheme. Following the standardisation of the narrowband AMR codec, the wideband AMR scheme referred to as the AMR-WB arrangement and encoding the 0 - 7 KHz band was also developed, which will also be characterised in the book. Finally, the most recent AMR codec, namely the so-called AMR-WB+ scheme will also be the subject of our discussions.

Rcent research on sub-2.4 kbps speech codecs is also covered extensively in the book, where the aspects of auditory masking become more dominant. Finally, since the classic G.722 subband-ADPCM based wideband codec has become obsolete in the light of exciting new developments in compression, the most recent trend is to consider wideband speech and audio codecs, providing susbtantially enhanced speech quality. Motivated by early seminal work on transform-domain or frequency-domain based compression by Noll and his colleagues, in this field the wideband G.721.1 codec - which can be programmed to operate between 10 kbps and 32 kbps and hence lends itself to employment in HSDPA-style near-instantaneously adaptive wireless communicators - is the most attractive candidate. This codec is portrayed in the context of a sophisticated burst-by-burst adaptive wideband turbo-coded Orthogonal Frequency Division Multiplex (OFDM) IMT in the book. This scheme is also capable of transmitting high-quality audio signals, behaving essentially as a high-quality waveform codec.

Mile-stones in Speech Coding History

Over the years a range of excellent monographs and text books have been published, characterising the state-of-the-art at its various stages of development and constituting significant mile-stones. The first major development in the history of speech compression can be considered the invention of the vocoder, dating back to as early as 1939. Delta modulation was contrived in 1952 and later it became well established following Steele's monograph on the

topic in 1975 [3]. Pulse Coded Modulation (PCM) was first documented in detail in Cattermole's classic contribution in 1969 [4]. However, it was realised in 1967 that predictive coding provides advantages over memory-less coding techniques, such as PCM. Predictive techniques were analysed in depth by Markel and Gray in their 1976 classic treatise [5]. This was shortly followed by the often cited reference [6] by Rabiner and Schafer. Also Lindblom and Ohman contributed a book in 1979 on speech communication research [7].

The foundations of auditory theory were layed down as early as 1970 by Tobias [8], but these principles were not exploited to their full potential until the invention of the analysis by synthesis (AbS) codecs, which were heralded by Atal's multi-pulse excited codec in the early eighties [9]. The waveform coding of speech and video signals has been comprehensively documented by Jayant and Noll in their 1984 monograph [10]. During the eighties the speech codec developments were fuelled by the emergence of mobile radio systems, where spectrum was a scarce resource, potentially doubling the number of subscribers and hence the revenue, if the bitrate could be halved.

The RPE principle - as a relatively low-complexity analysis by synthesis technique - was proposed by Kroon, Deprettere and Sluyter in 1986 [11], which was followed by further research conducted by Vary [12, 13] and his colleagues at PKI in Germany and IBM in France, leading to the 13 kbps Pan-European GSM codec. This was the first standardised AbS speech codec, which also employed long-term prediction (LTP), recognising the important role the pitch determination plays in efficient speech compression [14, 15]. It was in this era, when Atal and Schroeder invented the Code Excited Linear Predictive (CELP) principle [16], leading to perhaps the most productive period in the history of speech coding during the eighties. Some of these developments were also summarised for example by O'Shaughnessy [17], Papamichalis [18], Deller, Proakis and Hansen [19].

It was during this era that the importance of speech perception and acoustic phonetics [20] was duly recognised for example in the monograph by Lieberman and Blumstein. A range of associated speech quality measures were summarised by Quackenbush, Barnwell III and Clements [21]. Nearly concomitantly Furui also published a book related to speech processing [22]. This period witnessed the appearance of many of the speech codecs seen in Figure 1, which found applications in the emerging global mobile radio systems, such as IS-54, JDC, etc. These codecs were typically associated with source-sensitivity matched error protection, where for example Steele, Sundberg and Wong [23–26] have provided early insights on the topic. Further sophisticated solutions were suggested for example by Hagenauer [27].

Both the narrow-band and wide-band AMR, as wello as the AMR-WB+ (AMR) codecs [28, 29] are capable of adaptively adjusting their bitrate. This also allows the user to adjust the ratio between the speech bit rate and the channel coding bit rate constituting the error protection oriented redundancy according to the prevalent near-instantaneous channel conditions in HSDPA-style transceivers. When the channel quality is inferior, the speech encoder operates at low bit rates, thus accommodating more powerful forward error control within the total bit rate budget. By contrast, under high-quality channel conditions the speech encoder may benefit from using the total bit rate budget, yielding high speech quality, since in this high-rate case low redundancy error protection is sufficient. Thus, the AMR concept allows the system to operate in an error-resilient mode under poor channel conditions, while benefitting from a better speech quality under good channel conditions. Hence, the source coding scheme must be designed for seamless switching between rates available without annoying artifacts.

Overview of MPEG-4 Audio

The Moving Picture Experts Group (MPEG) was first established by the International Standard Organisation (ISO) in 1988 with the aim of developing a full audio-visual coding standard referred to as MPEG-1 [30–32]. The audio-related section MPEG-1 was designed to encode digital stereo sound at a total bit rate of 1.4 to 1.5 Mbps - depending on the sampling frequency, which was 44.1 kHz or 48 kHz - down to a few hundred kilobits per second [33]. The MPEG-1 standard is structured in layers, from Layer I to III. The higher layers achieve a higher compression ratio, albeit at an increased complexity. Layer I achieves perceptual transparency, i.e. subjective equivalence with the uncompressed original audio signal at 384 kbit/s, while Layer II and III achieve a similar subjective quality at 256 kbit/s and 192 kbit/s, respectively [34–38].

MPEG-1 was approved in November 1992 and its Layer I and II versions were immediately employed in practical systems. However, the MPEG Audio Layer III, MP3 for short only became a practical reality a few years later, when multimedia PCs were introduced having improved processing capabilities and the emerging Internet sparked off a proliferation of MP3 compressed teletraffic. This changed the face of the music world and its distribution of music. The MPEG-2 backward compatible audio standard was approved in 1994 [39], providing an improved technology that would allow those who had already launched MPEG-1 stereo audio services to upgrade their system to multichannel mode, optionally also supporting a higher number of channels at a higher compression ratio. Potential applications of the multichannel mode are in the field of quadraphonic music distribution or cinemas. Furthermore, lower sampling frequencies were also incorporated, which include 16, 22.05, 24, 32, 44.1 and 48 kHz [39]. Concurrently, MPEG commenced research into even higher-compression schemes, relinquishing the backward compatibility requirement, which resulted in the MPEG-2 Advanced Audio Coding standard (AAC) standard in 1997 [40]. This provides those who are not constrained by legacy systems to benefit from an improved multichannel coding scheme. In conjunction with AAC, it is possible to achieve perceptual transparent stereo quality at 128 kbit/s and transparent multichannel quality at 320 kbit/s for example in cinema-type applications.

The MPEG-4 audio recommendation is the latest standard completed in 1999 [41–45], which offers in addition to compression further unique features that will allow users to interact with the information content at a significant higher level of sophistication than is possible today. In terms of compression, MPEG-4 supports the encoding of speech signals at bit rates from 2 kbit/s up to 24 kbit/s. For coding of general audio, ranging from very low bit rates up to high quality, a wide range of bit rates and bandwidths are supported, ranging from a bit rate of 8 kbit/s and a bandwidth below 4 kHz to broadcast quality audio, including monoaural representations up to multichannel configuration.

The MPEG-4 audio codec includes coding tools from several different encoding families, covering parametric speech coding, CELP-based speech coding and Time/Frequency (T/F) audio coding, which are characterised in Figure 11.1. It can be observed that a parametric coding scheme, namely Harmonic Vector eXcitation Coding (HVXC) was selected for covering the bit rate range from 2 to 4 kbit/s. For bit rates between 4 and 24 kbit/s, a CELP-coding scheme was chosen for encoding narrowband and wideband speech signals. For encoding general audio signals at bit rates between 8 and 64 kbit/s, a time/frequency coding scheme based on the MPEG-2 AAC standard [40] endowed with additional tools is used. Here, a com-

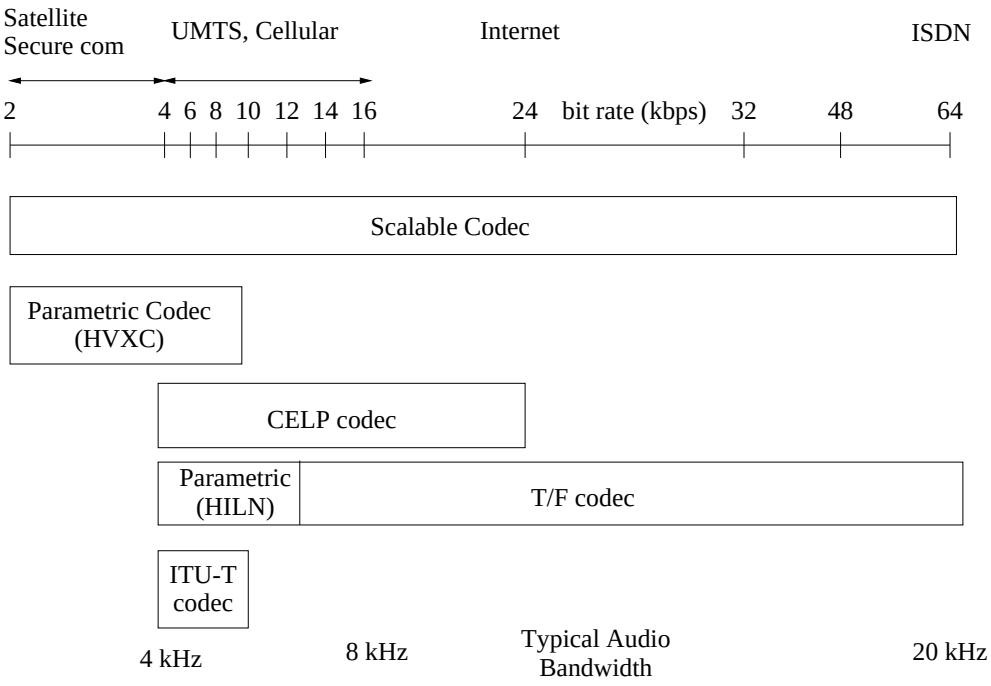

Figure 2: MPEG-4 framework [41].

bination of different techniques was established, because it was found that maintaining the required performance for representing speech and music signals at all desired bit rates cannot be achieved by selecting a single coding architecture. A major objective of the MPEG-4 audio encoder is to reduce the bit rate, while maintaining a sufficiently high flexibility in terms of bit rate selection. The MPEG-4 codec also offers other new functionalities, which include bit rate scalability, object-based of a specific audio passage for example, played by a certain instrument representation, robustness against transmission errors and supporting special audio effects.

MPEG-4 consists of Versions 1 and 2. Version 1 [41] contains the main body of the standard, while Version 2 [46] provides further enhancement tools and functionalities, that includes the issues of increasing the robustness against transmission errors and error protection, low-delay audio coding, finely grained bit rate scalability using the Bit-Sliced Arithmetic Coding (BSAC) tool, the employment of parametric audio coding, using the CELP-based silence compression tool and the 4 kbit/s extended variable bit rate mode of the HVXC tool. Due to the vast amount of information contained in the MPEG-4 standard, we will only consider some of its audio compression components, which include the coding of natural speech and audio signals. Readers who are specifically interested in text-to-speech synthesis or synthetic audio issues are referred to the MPEG-4 standard [41] and to the contributions by Scheirer *et al.* [47, 48] for further information. Most of the material in this chapter will be based on an amalgam of References [34–38, 40, 41, 43, 44, 46, 49]. In the next few sections,

the operations of each component of the MPEG-4 audio component will be highlighted in greater detail. As an application example, we will employ the Transform-domain Weighted Interleaved Vector Quantization (TWINVQ) coding tool, which is one of the MPEG-4 audio codecs in the context of a wireless audio transceiver in conjunction with space-time coding [50] and various Quadrature Amplitude Modulation (QAM) schemes [51]. The audio transceiver is introduced in Section 11.5 and its performance is discussed in Section 11.5.6.

Motivation and Outline of the Book

During the early 1990s Atal, Cuperman and Gersho [52] have edited prestigious contributions on speech compression. Also Ince [53] contributed a book in 1992 related to the topic. Anderson and Mohan co-authored a monograph on source and channel coding in 1993 [54]. Research-oriented developments were then consolidated in Kondoz' excellent monograph in 1994 [55] and in the multi-authored contribution edited by Keijn and Paliwal [56] in 1995. The most recent addition to the above range of contributions is the second edition of O'Shaughnessy well-referenced book cited above. However, at the time of writing no book spans the entire history of speech and audio compression, which is the goal of this volume.

Against this backcloth, this book endeavours to review the recent history of speech compression and communications in the era of wireless turbo-transceivers and joint source/channel coding. We attempt to provide the reader with a historical perspective, commencing with a rudimentary introduction to communications aspects, since throughout the book we illustrate the expected performance of the various speech codecs studied also in the context of jointly optimized wireless transceivers.

The book is constituted by four parts. Part I and II are covering classic background material on speech signals, predictive waveform codecs and analysis-by-synthesis codecs as well as on the entire speech and audio coding standardisation scene. The bulk of the book is constituted by the research-oriented Part III and IV, covering both standardised and proprietary speech codecs - including the most recent AMR-WB+ and the MPEG-4 audio codecs, as well as cutting-edge wireless turbo transceivers.

Specifically, **Chapters 1 and 2 of Part I** provide a rudimentary introduction to speech signals, classic waveform coding as well as predictive coding, respectively, quantifying the overall performance of the various speech codecs, in order to render our treatment of the topics as self-contained and all-encompassing as possible.

Part II of the book is centred around analysis by synthesis based coding, reviewing the classic principles in Chapter 3 as well as both narrow and wideband spectral envelope quantisation in Chapter 4. RPE and CELP coding are the topic of Chapters 5 and 6, which are followed by a detailed chapter on the entire plethora of existing forward-adaptive standardised CELP codecs in Chapter 7 and on their associated source-sensitivity matched channel coding schemes. The subject of Chapter 8 is both proprietary and standard backward-adaptive CELP codecs, which is concluded with a system design example based on a low-delay, multi-mode wireless transceiver.

The research-oriented **Part III** of the book are dedicated to a range of standard and proprietary wideband coding techniques and wireless systems. As an introduction to the wideband coding scene, in Chapter 9 the classic subband-based G.722 wideband codec is reviewed first, leading to the discussion of numerous low-rate wideband voice and audio codecs. Chapter 9

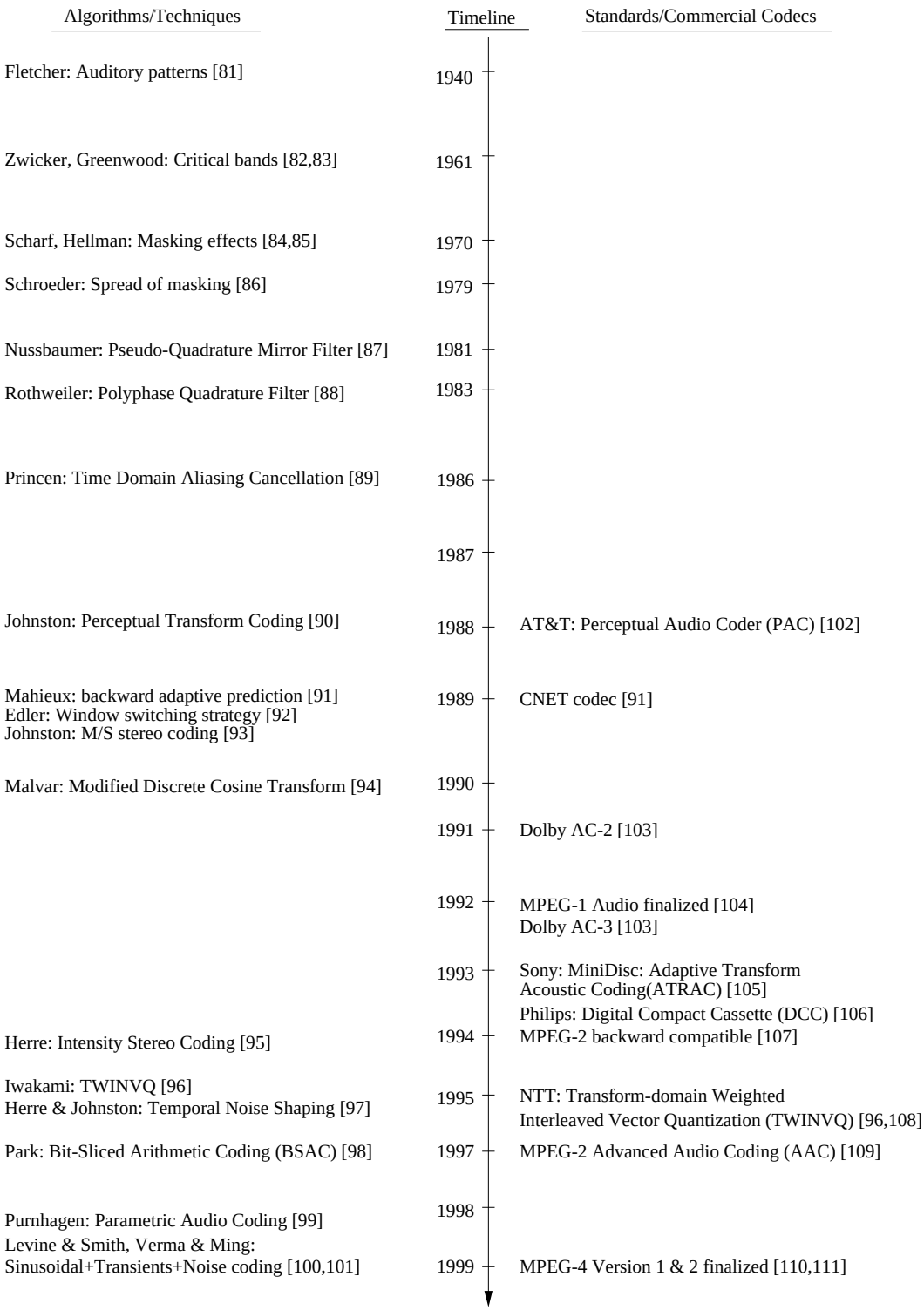

Figure 3: Important milestones in the development of perceptual audio coding.

also contains diverse sophisticated wireless voice- and audio-system design examples, including a turbo-coded Orthogonal Frequency Division Multiplex (OFDM) wideband audio system design study. This is followed by a wideband voice transceiver application example using the AMR-WB codec, a source-sensitivity matched Irregular Convolutional Code (IRCC) and Extrinsic Information Transfer (EXIT) charts for achieving a near-capacity system performance. Chapter 9 is concluded with the protrayal of the AMR-WB+ codec. In Chapter 10 of **Part III** we detailed the principles behind the MPEG-4 codec and comparatively studied the performance of the MPEG-4 and AMR-WB audio/speech codecs combined with various sophisticated wireless transceivers. Amongst others, a jointly optimised source-coding, outer unequal protection Non-Systematic Convolutional (NSC) channel-coding, inner Trellis Coded Modulation (TCM) and spatial diversity aided Space-Time Trellis Coded (STTC) turbo transceiver was investigated. The employment of TCM provided further error protection without expanding the bandwidth of the system and by utilising STTC spatial diversity was attained, which rendered the error statistics experienced pseudo-random, as required by the TCM scheme, since it was designed for Gaussian channels inflicting randomly dispersed channel errors. Finally, the performance of the STTC-TCM-2NSC scheme was enhanced with the advent of an efficient iterative joint decoding structure.

Chapters 11-17 of **Part IV** are all dedicated to sub-4kbps codecs and their wireless transceivers, while Chapter 18 is devoted to speech quality evaluation techniques as well as to a rudimentary comparison of various speech codecs and transceivers. The last chapter of the book is on VoIP.

This book is naturally limited in terms of its coverage of these aspects, simply owing to space limitations. We endeavoured, however, to provide the reader with a broad range of applications examples, which are pertinent to a range of typical wireless transmission scenarios.

Our hope is that the book offers you - the reader - a range of interesting topics, portraying the current state-of-the-art in the associated enabling technologies. In simple terms, finding a specific solution to a voice communications problem has to be based on a compromise in terms of the inherently contradictory constraints of speech quality, bitrate, delay, robustness against channel errors, and the associated implementational complexity. Analysing these trade-offs and proposing a range of attractive solutions to various voice communications problems is the basic aim of this book.

Again, it is our hope that the book underlines the range of contradictory system design trade-offs in an unbiassed fashion and that you will be able to glean information from it, in order to solve your own particular wireless voice communications problem, but most of all that you will find it an enjoyable and relatively effortless reading, providing you - the reader - with intellectual stimulation.

Lajos Hanzo
Clare Somerville
Jason Woodard

Acknowledgements

The book has been conceived in the Electronics and Computer Science Department at the University of Southampton, although Dr. Somerville and Dr. Woodard have moved on in the mean-time. We are indebted to our many colleagues who have enhanced our understanding of the subject, in particular to Prof. Emeritus Raymond Steele. These colleagues and valued friends, too numerous all to be mentioned, have influenced our views concerning various aspects of wireless multimedia communications and we thank them for the enlightment gained from our collaborations on various projects, papers and books. We are grateful to Jan Brecht, Jon Blogh, Marco Breiling, Marco del Buono, Sheng Chen, Stanley Chia, Byoung Jo Choi, Joseph Cheung, Peter Fortune, Sheyam Domeya, Lim Dongmin, Dirk Didascalou, Stephan Ernst, Eddie Green, David Greenwood, Hee Thong How, Thomas Keller, Ee-Lin Kuan, Joerg Kliewer, W.H. Lam, C.C. Lee, M.A. Nofal, Xiao Lin, Chee Siong Lee, Tong-Hooi Liew, Soon-Xin Ng, Matthias Muenster, Noor Othman, Vincent Roger-Marchart, Redwan Salami, David Stewart, Jeff Torrance, Spiros Vlahoyiannatos, Jin Wang, William Webb, John Williams, Jason Woodard, Choong Hin Wong, Henry Wong, James Wong, Lie-Liang Yang, Bee-Leong Yeap, Mong-Suan Yee, Kai Yen, Andy Yuen and many others with whom we enjoyed an association.

We also acknowledge our valuable associations with the Virtual Centre of Excellence in Mobile Communications, in particular with its Chief Executives, Dr. Tony Warwick and Dr. Walter Tuttlebee, Dr. Keith Baughan and other members of its Executive Committee, Professors Hamid Aghvami, Mark Beach, John Dunlop, Barry Evans, Joe McGeehan, Steve MacLaughlin, Rahim Tafazolli. Our sincere thanks are also due to John Hand and Nafeesa Simjee, the EPSRC, UK; Dr. Joao Da Silva, Dr Jorge Pereira, Bartholome Arroyo, Bernard Barani, Demosthenes Ikonomou and other colleagues from the Commission of the European Communities, Brussels, Belgium; Andy Wilton, Luis Lopes and Paul Crichton from Motorola ECID, Swindon, UK for sponsoring some of our recent research.

We feel particularly indebted to Hee Thong How for his invaluable contributions to the book by coauthoring some of the chapters and to Rita Hanzo as well as Denise Harvey for their skillful assistance in typesetting the manuscript in Latex. Similarly, our sincere thanks are due to Mark Hammond, Jennifer Beal, Sarah Hinton and a number of other staff from John Wiley & Sons for their kind assistance throughout the preparation of the camera-ready manuscript. Finally, our sincere gratitude is due to the numerous authors listed in the Author Index - as well as to those, whose work was not cited due to space limitations - for their contributions to the state-of-the-art, without whom this book would not have materialised.

Lajos Hanzo
Clare Somerville
Jason Woodard

Part I

Speech Signals and Waveform Coding

Part III

Wideband Coding and Transmission

Chapter 10

Advanced Multi-Rate Speech Transceivers

H-T. How and L. Hanzo

10.1 Introduction

[1] Recent speech coding research efforts have been successful in creating a range of both narrow- and wide-band multimode and multirate coding schemes, many of which have found their way into standardised codecs, such as the Advanced Multi-Rate (AMR) codec and its wideband version known as the AMR-WB scheme proposed for employment in the third-generation wireless systems. Other multimode solutions have been used in the MPEG-4 codec, which will be investigated in the next chapter. In multimode coding schemes [335, 336], a mode selection process is invoked and the specific coding mode best suited to the local character of the speech signal is selected from a predetermined set of modes. This technique dynamically tailors the coding scheme to the widely varying local acoustic-phonetic character of the speech signal.

Multi-rate coding on the other hand facilitates the assignment of a time-variant number of bits for a frame, adapting the encoding rate on the basis of the local phonetic character of the speech signal or the network conditions. This is particularly useful in digital cellular communications, where one of the major challenges is that of designing an encoder that is capable of providing high quality speech for a wide variety of channel conditions. Ideally, a good solution must provide the highest possible speech quality under perfect channel conditions, while maintaining an error-resilient behaviour in hostile channel environments. Traditionally, existing digital cellular applications have employed a single coding mode where a fixed source/channel bit allocation provides a compromise solution between the perfect and hostile

[1] This chapter is based on H.T. How, T.H. Liew, E.L. Kuan, L-L. Yang and L. Hanzo: A Redundant Residue Number System Coded Burst-by-Burst Adaptive Joint-Detection Based CDMA Speech Transceiver, IEEE Transactions on Vehicular Technology, Volume 55, Issue 1, Jan. 2006, pp 387 - 397

channel conditions. Clearly, a coding solution which is well suited for high-quality channels would use most of the available bits for source coding in conjunction with only minimal error protection, while a solution designed for poor channels would use a lower rate speech encoder along with more powerful forward error protection. Due to the powerful combination of channel equalization, interleaving and channel coding, near-error-free transmission can be achieved down to a certain threshold of the Carrier to Interferer ratio (C/I). However, below this threshold, the error correction code is likely to fail in removing the transmission errors, with the result that the residual errors may cause annoying artifacts in the reconstructed speech signal.

Therefore, in existing systems typically a worst case design is applied, where the channel coding scheme is sufficiently powerful to remove most transmission errors, as long as the system operates within a reasonable C/I range. However, the drawback of this solution is that the speech quality becomes lower than necessary under good channel conditions, since a high proportion of the gross bit rate is dedicated to channel coding.

The Adaptive Multi-Rate (AMR) concept [28] solves this 'resource allocation' problem in a more intelligent way. Specifically, the ratio between the speech bit rate and the error protection oriented redundancy is adaptively adjusted according to the prevalent channel conditions. While the channel quality is inferior, the speech encoder operates at low bit rates, thus accommodating powerful forward error control within the total bit rate budget. By contrast, under high channel conditions the speech encoder may benefit from using the total bit rate budget, yielding high speech quality, since in this high-rate case low redundancy error protection is sufficient. Thus, the AMR concept allows the system to operate in an error-resilient mode under poor channel conditions, while benefitting from a better speech quality under good channel conditions. This is achieved by dynamically splitting the gross bit rate of the transmission system between source and channel coding according to the instantaneous channel conditions. Hence, the source coding scheme must be designed for seamless switching between rates available without annoying artifacts.

In this chapter, we first give an overview of the AMR narrowband codec [29], which has been standardised by ETSI [28, 29]. The AMR codec is capable of operating in both the full-rate and half-rate speech traffic channels of GSM. It is also amenable to adapting the source coding and channel coding bit rates according to the quality of the radio channel. As stated above, most speech codecs employed in communication systems - such as for example the existing GSM speech codecs (full-rate [362], half-rate [363] and enhanced full-rate [364]) - operate at a fixed bit rate, with a trade-off between source coding and channel coding. However, estimating the channel quality and adjusting the transceiver's bit rate adaptively according to the channel conditions has the potential of improving the system's error resilience and hence the speech quality experienced over high error-rate wireless channels.

The inclusion of an AMR Wideband (AMR-WB) mode has also been under discussion, with feasibility studies [337, 338] being conducted at the time of writing for applications in GSM networks, as well as for the evolving Third Generation (3G) systems [339]. With aim of providing a system-design example for such intelligent systems, during our forthcoming discourse in this chapter we will characterize the error sensitivity of the AMR encoder's output bits so that the matching channel encoder can be carefully designed for providing the required protection for the speech bits, which are most sensitive to transmission errors. The proposed intelligent adaptive multirate voice communications system will be described in Section 10.4.

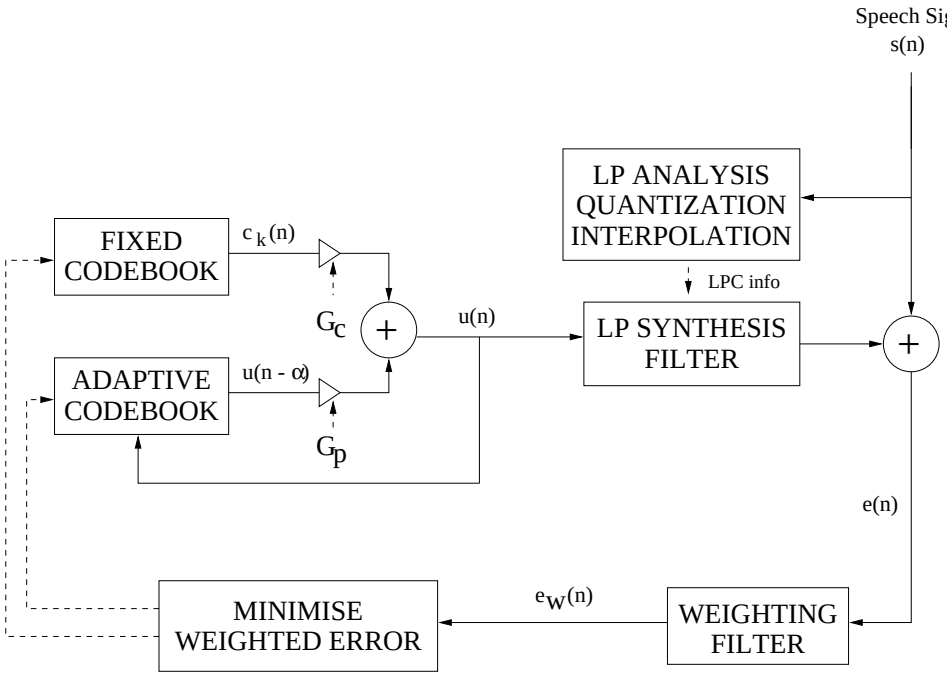

Figure 10.1: Schematic of ACELP speech encoder.

10.2 The Adaptive Multi-Rate Speech Codec

10.2.1 Overview

The AMR codec employs the Algebraic Code-Excited Linear Predictive (ACELP) model [365, 366] shown in Figure 10.1. Here we provide a brief overview of the AMR codec following the approach of [28, 29, 257]. The AMR codec's complexity is relatively low and hence it can be implemented cost-efficiently. This codec operates on a 20ms frame of 160 speech samples, and generates encoded blocks of 95, 103, 118, 134, 148, 159, 204 and 244 bits/20ms. This leads to bit rates of 4.75, 5.15, 5.9, 6.7, 7.4, 7.95, 10.2 and 12.2 kbit/s, respectively. Explicitly, the AMR speech codec provides eight different modes and their respective Segmental SNR performance was shown in Figure 10.2.

Multirate coding [335] supports a variable allocation of bits for a speech frame, adapting the rate to the instantaneous local phonetic character of the speech signal, to the channel quality or to network conditions. This is particularly useful in digital cellular communications, where one of the major challenges is that of designing a codec that is capable of providing high quality speech for a wide variety of channel conditions. Ideally, a good solution must provide the highest possible quality under perfect channel conditions, unimpaired by the channel, while also maintaining good quality in hostile high error-rate channel environments. The codec mode adaptation is a key feature of the new AMR standard that has not been used in any prior mobile standard. At a given fixed gross bit rate, this mechanism of adapting the source coding rate has the potential of altering the partitioning between the speech source bit

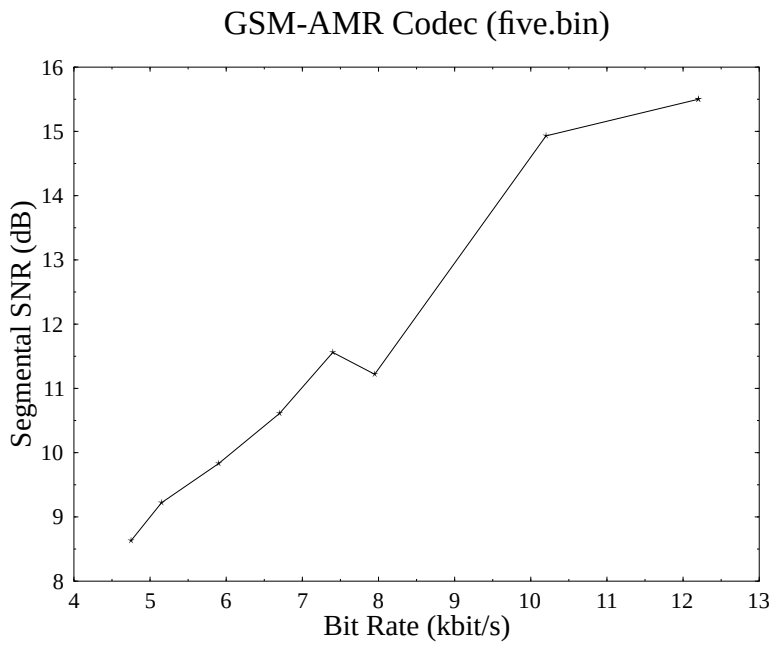

Figure 10.2: Segmental SNR performance of the AMR codec, operating at bit rates in the range between 4.75 kbit/s and 12.2 kbit/s.

rate and the redundancy added for error protection. Hence, the AMR codec will be invoked in our Burst-by-Burst Adaptive Quadrature Amplitude Modulation Code Division Multiple Access (BbB-AQAM/CDMA) transceiver.

As shown in Figure 10.1, the Algebraic Code Excited Linear Prediction (ACELP) encoder operates on the sampled input speech signal $s(n)$ and Linear Prediction Coding(LPC) is applied to each speech segment. The coefficients of this predictor are used for constructing an LPC synthesis filter $1/(1 - A(z))$, which describes the spectral envelope of the speech segment [335, 367]. An Analysis-by-Synthesis (AbS) procedure is employed, in order to find the particular excitation that minimizes the weighted Minimum Mean Square Error (MMSE) between the reconstructed and original speech signal. The weighting filter is derived from the LPC synthesis filter and takes into account the psychoacoustic quantisation noise masking effect, namely that the quantization noise in the spectral neighbourhood of the spectrally prominent speech formants is less perceptible [335, 367]. In order to reduce the complexity, the adaptive and fixed excitation codebooks are searched sequentially in order to find the perceptually best codebook entry, first for the adaptive codebook contribution, and then for the fixed codebook entry. The adaptive codebook consists of time-shifted versions of past excitation sequences and describes the long-term characteristics of the speech signal [335, 367].

Three of the AMR coding modes correspond to existing standards, which renders communication systems employing the new AMR codec interoperable with other systems. Specifi-

cally, the 12.2 kbit/s mode is identical to the GSM Enhanced Full Rate (EFR) standard [339], the 12.2 and 7.4 kbit/s modes [368] correspond to the US1 and EFR (IS-641) codecs of the TDMA (IS-136) system, and the 6.7 kbit/s mode is equivalent to the EFR codec of the Japanese PDC system [335]. For each of the codec modes, there exist corresponding channel codecs, which perform the mapping between the speech source bits and the fixed number of channel coded bits.

In the forthcoming subsections, we will give a functional description of the AMR codec's operation in the 4.75 and 10.2 kbit/s modes. These two bit rates will be used in our investigations in order to construct a dual-mode speech transceiver in Section 10.4.

10.2.2 Linear Prediction Analysis

A 10th order LPC analysis filter is employed for modelling the short term correlation of the speech signal $s(n)$. Short-term prediction, or linear predictive analysis is performed once for each 20ms speech frame using the Levinson-Durbin algorithm [367]. The LP coefficients are transformed to the Line Spectrum Frequencies (LSF) for quantization and interpolation. The employment of the LSF [369] representation for quantization of the LPC coefficients, is motivated by their advantageous statistical properties. Specifically, within each speech frame, there is a strong intra-frame correlation due to the ordering property of the neighbouring LSF values [367]. This essentially motivates the employment of vector quantization. The interpolated quantized and unquantized LSFs are converted back to the LP filter coefficients, in order to construct the synthesis and weighting filters at each subframe. The synthesis filter shown in Figure 10.1 is used in the decoder for producing the reconstructed speech signal from the received excitation signal $u(n)$.

10.2.3 LSF Quantization

In the AMR codec, the LSFs are quantized using interframe LSF prediction and Split Vector Quantization (SVQ) [28]. The SVQ aims to split the 10-dimensional LSF vector into a number of reduced-dimension LSF subvectors, which simplifies the associated codebook entry matching and search complexity. Specifically, the proposed configuration minimizes the average Spectral Distortion (SD) [370] achievable at a given total complexity. Predictive vector quantization is used [28] and the 10-component LSF vectors are split into 3 LSF subvectors of dimension 3, 3 and 4. The bit allocations for the three subvectors will be described in Section 10.2.7 for the 4.75- and 10.2 kbit/s speech coding modes.

10.2.4 Pitch Analysis

Pitch analysis using the adaptive codebook approach models the long-term periodicity, i.e., the pitch of the speech signal. It produces an output, which is an amplitude- scaled version of the adaptive codebook of Figure 10.1 based on previous excitations. The excitation signal $u(n) = G_p u(n-\alpha) + G_c c_k(n)$ seen in Figure 10.1 is determined from its G_p-scaled history after adding the G_c-scaled fixed algebraic codebook vector c_k for every 5ms subframe. The optimum excitation is chosen on the basis of minimising the mean squared error E_w over the subframe.

Subframe	Subset	Pulse: Positions
1	1	i_0: 0,5,10,15,20,25,30,35
		i_1: 2,7,12,17,22,27,32,37
	2	i_0: 1,6,11,16,21,26,31,36
		i_1: 3,8,13,18,23,28,33,38
2	1	i_0: 0,5,10,15,20,25,30,35
		i_1: 3,8,13,18,23,28,33,38
	2	i_0: 2,7,12,17,22,27,32,37
		i_1: 4,9,14,19,24,29,34,39
3	1	i_0: 0,5,10,15,20,25,30,35
		i_1: 2,7,12,17,22,27,32,37
	2	i_0: 1,6,11,16,21,26,31,36
		i_1: 4,9,14,19,24,29,34,39
4	1	i_0: 0,5,10,15,20,25,30,35
		i_1: 3,8,13,18,23,28,33,38
	2	i_0: 1,6,11,16,21,26,31,36
		i_1: 4,9,14,19,24,29,34,39

Table 10.1: Pulse amplitudes and positions for 4.75 kbit/s AMR codec mode [28].

Track	Pulse	Positions
1	i_0, i_4	0,4,8,12,16,20,24,28,32,36
2	i_1, i_5	1,5,9,13,17,21,25,29,33,37
3	i_2, i_6	2,6,10,14,18,22,26,30,34,38
4	i_3, i_7	3,7,11,15,19,23,27,31,35,39

Table 10.2: Pulse amplitudes and positions for 10.2 kbit/s AMR codec code [28].

In an optimal codec, the fixed codebook index and codebook gain as well as the adaptive codebook parameters would all be jointly optimized in order to minimize E_w [371]. However, in practice this is unfeasible due to the associated excessive complexity. Hence, a sequential sub-optimal approach is applied in the AMR codec, where the adaptive codebook parameters are determined first under the assumption of zero fixed codebook excitation component, i.e., $G_c = 0$, since at this optimisation stage no fixed codebook entry was determined. Then, given that the adaptive codebook parameters are found, which consist of the delay and gain of the pitch filter, the fixed codebook parameters are determined.

Most CELP codecs employ both so-called open-loop and closed-loop estimation of the adaptive codebook delay parameters, as is the case in the AMR codec. The open-loop estimate of the pitch period is used to narrow down the range of the possible adaptive codebook delay values and then the full closed-loop analysis-by-synthesis procedure is used for finding a high-resolution delay around the approximate open-loop position [366].

Parameter	1st subframe	2nd subframe	3rd subframe	4th subframe	Total per frame
LSFs					8+8+7=23 (1-23)
Pitch Delay	8 (24-31)	4 (49-52)	4 (62-65)	4 (83-86)	20
Fixed CB Index	9 (32-40)	9 (53-61)	9 (66-74)	9 (87-95)	36
Codebook Gains	8 (41-48)		8 (75-82)		16
Total					95/20ms=4.75 kbit/s
LSFs					8+9+9=26
Pitch Delay	8	5	8	5	26
Fixed CB Index	31	31	31	31	124
Codebook Gains	7	7	7	7	28
Total					204/20ms=10.2 kbit/s

Table 10.3: Bit allocation of the AMR speech codec at 4.75 kbit/s and 10.2 kbit/s [28]. The bit positions for 4.75 kbit/s mode, which are shown in round bracket assist in identifying the corresponding bits in Figure 10.5.

10.2.5 Fixed Codebook With Algebraic Structure

Once the adaptive codebook parameters are found, the fixed codebook is searched by taking into account the now known adaptive codebook vector. This sequential approach constitutes a trade-off between the best possible performance and the affordable computational complexity. The fixed codebook is searched by using an efficient non-exhaustive analysis-by-synthesis technique [372], minimizing the mean square error between the weighted input speech and the weighted synthesized speech.

The fixed, or algebraic codebook structure is specified in Table 10.1 and Table 10.2 for the 4.75 kbit/s and 10.2 kbit/s codec modes, respectively [28]. The algebraic fixed codebook structure is based on the so-called Interleaved Single-Pulse Permutation (ISPP) code design [371]. The computational complexity of the fixed codebook search is substantially reduced, when the codebook entries $c_k(n)$ used are mostly zeros. The algebraic structure of the excitation having only a few non-zero pulses allows for a fast search procedure. The non-zero elements of the codebook are equal to either +1 or -1, and their positions are restricted to the limited number of excitation pulse positions, as portrayed in Table 10.1 and 10.2 for the speech coding modes of 4.75 and 10.2 kbit/s, respectively.

More explicitly, in the 4.75 kbit/s codec mode, the excitation codebook contains 2 non-zero pulses, denoted by i_0 and i_1 in Table 10.1. Again, all pulses can have the amplitudes +1 or -1. The 40 positions in a subframe are divided into 4 so-called tracks. Two subsets of 2 tracks each are used for each subframe with one pulse in each track. Different subsets of tracks are used for each subframe, as shown in Table 10.1 and hence one bit is needed for encoding the subset used. The two pulse positions, i_0 and i_1 are encoded with the aid of 3 bits each, since both have eight legitimate positions in Table 10.1. Furthermore, the sign of each pulse is encoded using 1 bit. This gives a total of 1+2(3)+2(1)=9 bits for the algebraic excitation encoding in a subframe.

In the 10.2 kbit/s codec mode of Table 10.3 there are four tracks, each containing two pulses. Hence, the excitation vector contains a total of 4x2=8 non-zero pulses. All the pulses

can have the amplitudes of +1 or -1 and the excitation pulses are encoded using a total of 31 bits.

For the quantization of the fixed codebook gain, a gain predictor is used, in order to exploit the correlation between the fixed codebook gains in adjacent frames [28]. The fixed codebook gain is expressed as the product of the predicted gain based on previous fixed codebook energies and a correction factor. The correction factor is the parameter, which is coded together with the adaptive codebook gain for transmission over the channel. In the 4.75 kbit/s mode the adaptive codebook gains and the correction factors are jointly vector quantized for every 10 ms, while this process occurs every subframe of 5 ms in the 10.2 kbit/s mode.

10.2.6 Post-Processing

At the decoder, an adaptive postfilter [373] is used for improving the subjective quality of the reconstructed speech. The adaptive postfilter consists of a formant-based postfilter and a spectral tilt-compensation filter [373]. Adaptive Gain Control (AGC) is also used, in order to compensate for the energy difference between the synthesized speech signal, which is the output from the synthesis filter and the postfiltered speech signal.

10.2.7 The AMR Codec's Bit Allocation

The AMR speech codec's bit allocation is shown in Table 10.3 for the speech modes of 4.75 kbit/s and 10.2 kbit/s. For the 4.75 kbit/s speech mode, 23 bits are used for encoding the LSFs by employing split vector quantization. As stated before, the LSF vector is split into 3 subvectors of dimension 3, 3 and 4, and each subvector is quantized using 8, 8 and 7 bits, respectively. This gives a total of 23 bits for the LSF quantization of the 4.75 kbit/s codec mode.

The pitch delay is encoded using 8 bits in the first subframe and the relative delays of the other subframes are encoded using 4 bits. The adaptive codebook gain is quantized together with the above-mentioned correction factor of the fixed codebook gain for every 10ms using 8 bits. As a result, a total of 16 bits are used for encoding both the adaptive- and fixed codebook gains. As described in Section 10.2.5, 9 bits were used to encode the fixed codebook indices for every subframe, which resulted in a total of 36 bits per 20ms frame for the fixed codebook.

For the 10.2 kbit/s mode, the three LSF subvectors are quantized using 8, 9 and 9 bits respectively. This implies that 26 bits are used for quantizing the LSF vectors at 10.2 kbit/s, as shown in Table 10.3. The pitch delay is encoded using 8 bits in the first and third subframes and the relative delay of the other subframes is encoded using 5 bits. The adaptive codebook gain is quantized together with the correction factor of the fixed codebook gain using a 7-bit non-uniform vector quantization scheme for every 5ms subframe. The fixed codebook indices are encoded using 31 bits in each subframe, in order to give a total of 124 bits for a 20ms speech frame.

10.2.8 Codec Mode Switching Philosophy

In the AMR codec, the mode adaptation allows us to invoke a subset of at most 4 modes out of the 8 available modes [258]. This subset is referred to as the Active Codec Set (ACS). In the proposed BbB-AQAM/CDMA system the codec mode adaptation is based on the channel

quality, which is expressed as the MSE at the output of the multi-user CDMA detector [374]. The probability of switching from one mode to another is typically lower, than the probability of sustaining a specific mode.

Intuitively, frequent mode switching is undesirable due to the associated perceptual speech quality fluctuations. It is more desirable to have a mode selection mechanism that is primarily source-controlled, assisted by a channel-quality-controlled override. During good channel conditions, the mode switching process is governed by the local phonetic character of the speech signal and the codec will adapt itself to the speech signal characteristics in an attempt to deliver the highest possible speech quality. When the channel is hostile or the network is congested, transceiver control or external network control can take over the mode selection and allocate less bits to source coding, in order to increase the system's robustness or user capacity. By amalgamating the channel-quality motivated or network- and source-controlled processes, we arrive at a robust, high-quality system. Surprisingly, we found from our informal listening tests that the perceptual speech quality was not affected by the rate of codec mode switching, as it will be demonstrated in Section 10.8. This is due to the robust ACELP structure, whereby the main bit rate reduction is related to the fixed codebook indices, as shown in Table 10.3 for the codec modes of 4.75 kbit/s and 10.2 kbit/s.

As expected, the performance of the AMR speech codec is sensitive to transmission errors of the codec mode information. The corruption of the codec mode information that describes, which codec mode has to be used for decoding leads to complete speech frame losses, since the decoder is unable to apply the correct mode for decoding the received bit stream. Hence, robust channel coding is required in order to protect the codec mode information and the recommended transmission procedures were discussed for example by Bruhn *et al.* [257]. Furthermore, in transceiver-controlled scenarios the prompt transmission of the codec mode information is required for reacting to sudden changes of the channel conditions. In our investigations we assume that the signalling of the codec mode information is free from corruption, so that we can concentrate on other important aspects of the system.

Let us now briefly focus our attention on the robustness of the AMR codec against channel errors.

10.3 Speech Codec's Error Sensitivity

In this section, we will demonstrate that some bits are significantly more sensitive to channel errors than others, and hence these sensitive bits have to be better protected by the channel codec [371]. A commonly used approach in quantifying the sensitivity of a given bit is to invert this bit consistently in every speech frame and evaluate the associated Segmental SNR (SEGSNR) degradation. The error sensitivity of various bits of the AMR codec determined in this way is shown in Figure 10.3 for the bit rate of 4.75 kbit/s. Again, Figure 10.3 shows more explicitly the bit sensitivities in each speech subframe for the bit rate of 4.75 kbit/s, with the corresponding bit allocations shown in Table 10.3. For the sake of visual clarity, Subframe 4 (83-95) was not shown explicitly above, since it exhibited identical SEGSNR degradation to Subframe 2.

It can be observed from Figure 10.3 that the most sensitive bits are those of the LSF subvectors, seen at positions 1-23. The error sensitivity of the adaptive codebook delay is the highest in the first subframe, commencing at bit 24, as shown in Figure 10.3, which was

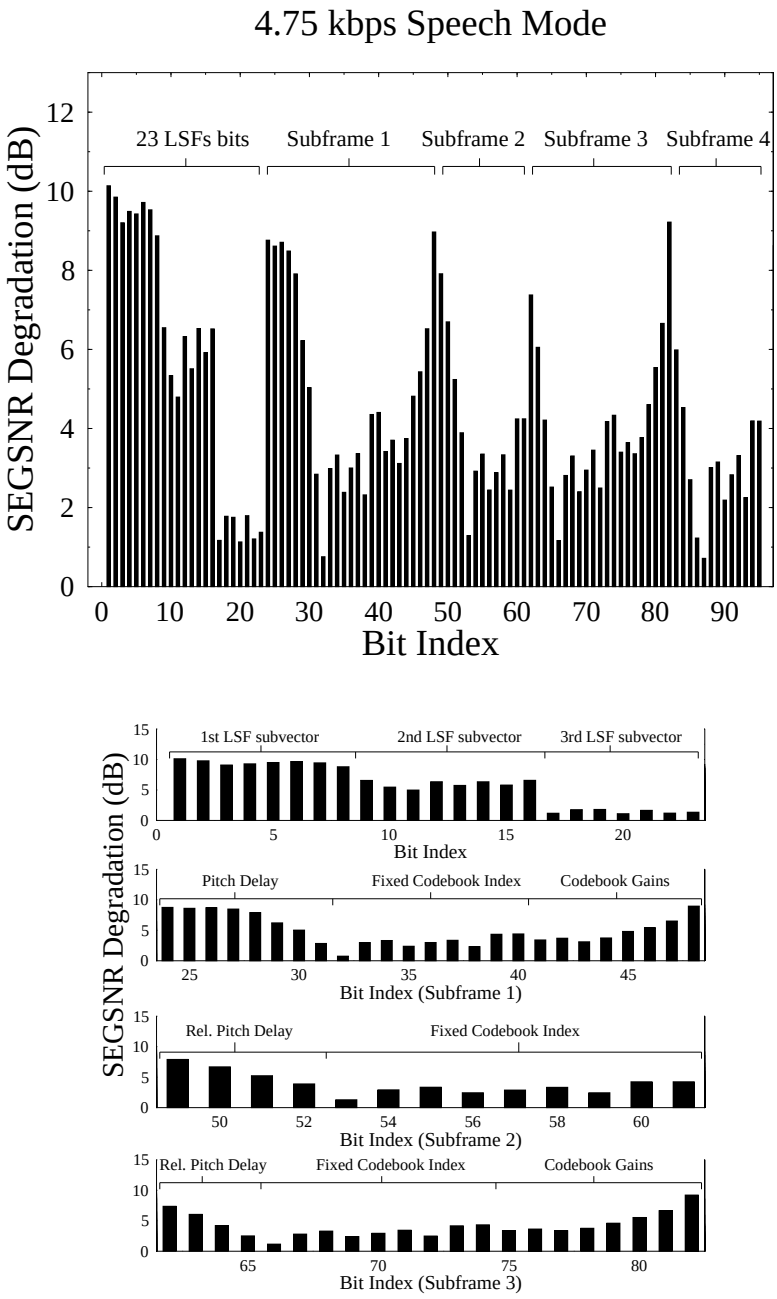

Figure 10.3: The SEGSNR degradations due to 100% bit error rate in the 95-bit, 20 ms AMR speech frame. The associated bit allocation can be seen in Table 10.3.

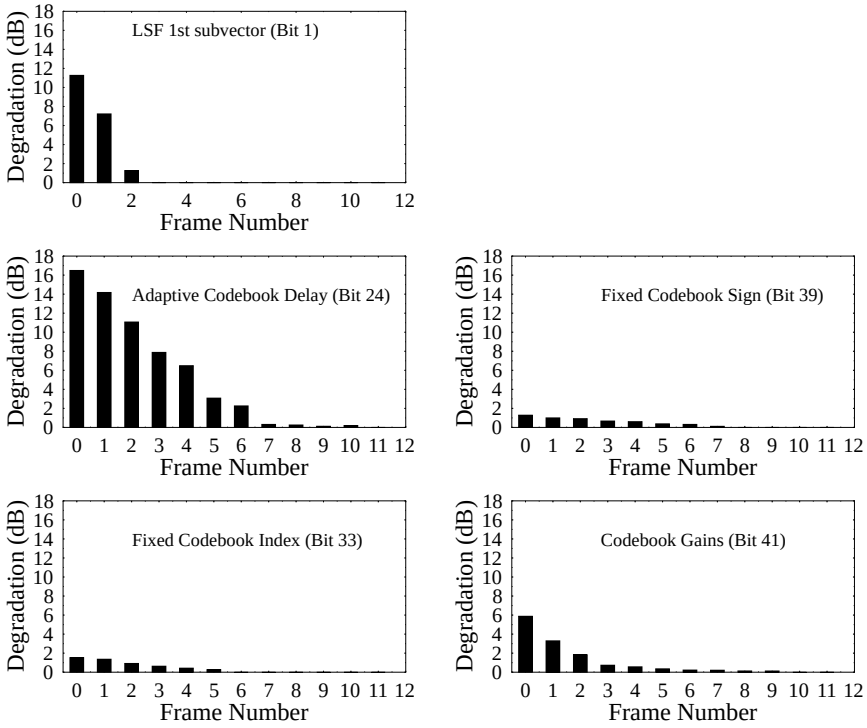

Figure 10.4: The SEGSNR degradation versus speech frame index for various bits.

encoded using 8 bits in Table 10.3. By contrast, the relative adaptive codebook delays in the next three subframes are encoded using 4 bits each, and a graceful degradation of the SEGSNR is observed in Figure 10.3. The next group of bits is constituted by the 8 codebook gains in decreasing order of bit sensitivity, as seen in Figure 10.3 for bit positions 41-48 of Subframe 1 and 75-82 of Subframe 3. The least sensitive bits are related to the fixed codebook pulse positions, which were shown for example at bit positions 54-61 in Figure 10.3. This is because, if one of the fixed codebook index bits is corrupted, the codebook entry selected at the decoder will differ from that used in the encoder only in the position of one of the non-zero excitation pulses. Therefore the corrupted codebook entry will be similar to the original one. Hence, the algebraic codebook structure used in the AMR codec is inherently quite robust to channel errors. The information obtained here will be used in Section 10.6.2 for designing the bit mapping procedure in order to assign the channel encoders according to the bit error sensitivities.

Although appealing in terms of its conceptual simplicity, the above approach we used for quantifying the error sensitivity of the various coded bits does not take into account the error propagation properties of different bits over consecutive speech frames. In order to obtain a better picture of the error propagation effects, we also employed a more elaborate error sensitivity measure [371]. Here, for each bit we find the average SEGSNR degradation due to a

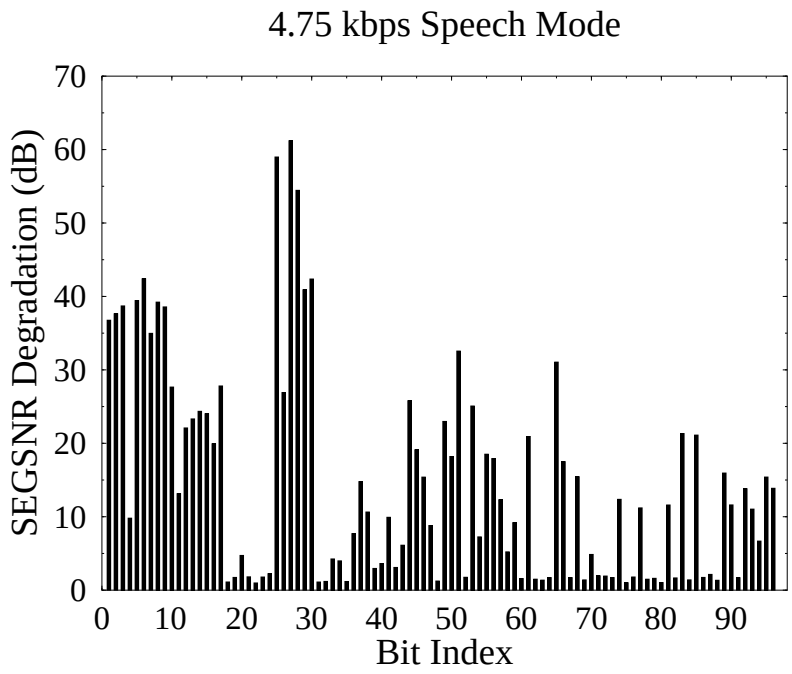

Figure 10.5: Average SEGSNR degradation due to single bit errors in various speech coded bits

single bit error both in the specific frame in which the error occurs and in consecutive frames. These effects are exemplified in Figure 10.4 for five different bits, where each of the bits belongs to a different speech codec parameter. More explicitly, Bit 1 represents the first bit of the first LSF subvector, which shows some error propagation effects due to the interpolation between the LSFs over consecutive frames. The associated SEGSNR degradation dies away over six frames. Bit 24 characterised in Figure 10.4 is one of the adaptive codebook delay bits and the corruption of this bit has the effect of a more prolonged SEGSNR degradation over 10 frames. The fixed codebook index bits of Table 10.3 are more robust and observed to be the least sensitive bits, as it was shown in Figure 10.3 earlier. This argument is supported by the example of Bit 33 in Figure 10.4, where a smaller degradation is observed over consecutive frames. A similar observation also applies to Bit 39 in Figure 10.4, which is the sign bit of the fixed codebook. By contrast, Bit 41 of the codebook gains produced a high and prolonged SEGSNR degradation profile.

We recomputed our bit-sensitivity results of Figure 10.3 using this second approach, in order to obtain Figure 10.5, taking into account the error propagation effects. More explicitly, these results were calculated by summing the SEGSNR degradations over all the frames, which were affected by the error. Again, these results are shown in Figure 10.5 and the associated bit positions can be identified with the aid of Table 10.3. The importance of the

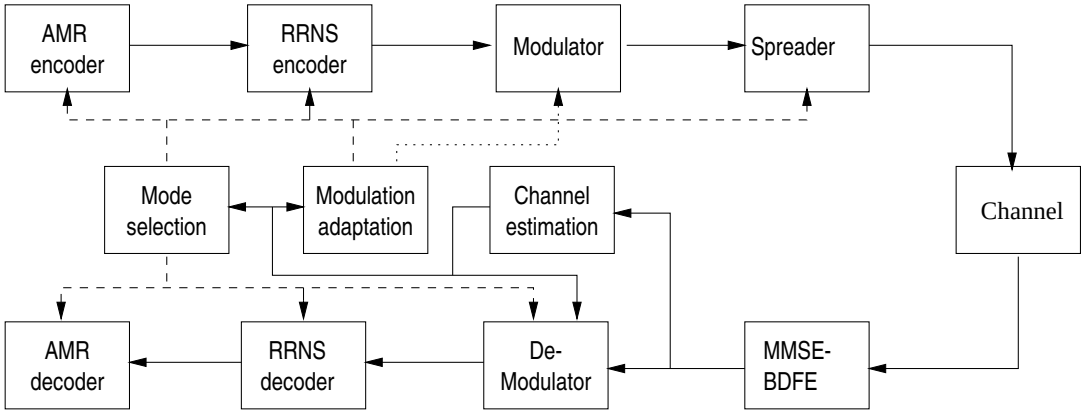

Figure 10.6: Schematic of the adaptive dual-mode JD-CDMA system

adaptive codebook delay bits became more explicit. By contrast, the significance of the LSFs was reduced, although still requiring strong error protection using channel coding.

Having characterised the error sensitivity of various speech bits, we will capitalise on this knowledge, in order to assign the speech bits to various bit protection classes, as it will be discussed in Section 10.6.2. Let us now consider the various components of our transceiver, which utilises the AMR codec in the next section. We will first discuss the motivation of employing multirate speech encoding in conjunction with a near-instantaneously adaptive transceiver, with a detailed background description of earlier contributions from various researchers.

10.4 System Background

The AMR concept is amenable to a range of intelligent configurations. When the instantaneous channel quality is low, the speech encoder operates at low bit rates, thus facilitating the employment of powerful forward error control within a fixed bit rate budget. By contrast, under favourable channel conditions the speech encoder may use its highest bit rate, implying high speech quality, since in this case weaker error protection is sufficient or a less robust, but higher bit rate transceiver mode can be invoked. However, the system must be designed for seamless switching between its operating rates without objectionable perceptual artifacts.

Das *et al.* provided an extensive review of multimode and multirate speech coding in [375]. Some of the earlier contributors in multimode speech coding included Taniguchi *et al.* [376], Kroon and Atal [377], Yong and Gersho [378], DeJaco *et al.* [379], Paksoy *et al.* [380] and Cellario *et al.* [381] . Further recent work on incorporating multirate speech coding into wireless systems was covered in a range of contributions [382]- [383]. Specifically, Yuen *et al.* [382] in their paper employed embedded and multimode speech codecs based on the Code Excited Linear Prediction (CELP) technique in combination with channel codecs using Rate Compatible Punctured Convolutional codes (RCPC) [384]. The combined speech and channel coding resulted in gross bit rates of 12.8 kbit/s and 9.6 kbit/s, supported by either

TDMA or CDMA multiple access techniques. The investigations showed that multimode CELP codecs performed better, than their embedded counterparts, and that adaptive schemes were superior to fixed-rate schemes.

LeBlanc *et al.* in [385] developed a low power, low delay, multirate codec suitable for indoor wireless communications. The speech codec was a modified version of the G.728 LD-CELP standard scheme [386], employing a multi-stage excitation configuration together with an adaptive codebook. A lower LPC predictor order of 10 was used, rather than 50 as in G.728, and a higher bandwidth expansion factor of 0.95, rather than 0.9883 was employed, which resulted in a more robust performance over hostile channels. This algorithm was investigated over indoor wireless channels assisted by 2-branch diversity, using QPSK modulation and wideband TDMA transmission. No channel coding was employed and the system's performance was not explicitly characterised in the paper. In [241], Kleider *et al.* proposed an adaptive speech transmission system utilising the Multi-Rate Sinusoidal Transform Codec (MRSTC), in conjunction with convolutional channel coding and Pulse Position Modulation (PPM). The MRSTC is based on the sinusoidal transform coding scheme proposed by McAulay [387]. The MRSTC was investigated further by the same authors for wireless and internet applications in [388], using a range of bit rates between 1.2 kbit/s and 9.6 kbit/s. The MRSTC was incorporated into a communication system employing convolutional channel coding and a fixed BPSK modulation scheme, and it was reported to give a nearly 9 dB in average spectral distortion reduction over the fixed-rate 9.6 kbit/s benchmarker.

In a contribution from the speech coding team at Qualcomm, Das *et al.* [389] illustrated using a multimode codec having four modes (Full-rate, Half-rate, Quarter-rate and Eight-rate), that the diverse characteristics of the speech segments can be adequately captured using variable rate codecs. It was shown that a reduced average rate can be obtained, achieving equivalent speech quality to that of a fixed full-rate codec. Specifically, a multimode codec with an average rate of 4 kbit/s achieved significantly higher speech quality than that of the equivalent fixed-rate codec. An excellent example of a recent standard variable-rate codec is the Enhanced Variable Rate Codec (EVRC), standardized by the Telecommunications Industry Association (TIA) as IS-127 [245]. This codec operates at a maximum rate of 8.5 kbit/s and at an average rate of about 4.1 kbit/s. The EVRC consists of three coding modes that are all based on the CELP model. The activation of one of the three modes is source-controlled, based on the estimation of the input signal state.

Multimode speech coding was also evaluated in an ATM-based environment by Beritelli *et al.* in [390]. The speech codec possessed seven coding rates, ranging from 0.4 to 16 kbit/s. Five different bit rates were allocated for voiced/unvoiced speech encoding, while two lower bit rates were generated for inactive speech periods, depending on the stationarity of the background noise. The variable-rate voice source was modelled using a Markov-model based process. The multimode coding scheme was compared to the 12 kbit/s CS-ACELP standard codec using the traditional ON-OFF voice generation model. It was found that the multimode codec performed better, than the CS-ACELP ON-OFF scheme, succeeding in minimizing the required transmission bandwidth by exploiting the near-instantaneous local characteristics of the speech waveform and it was also capable of synthesizing the background noise realistically.

Our discussion so far have been focused on source-controlled multirate codecs, where the coding algorithm responds to the time-varying local character of the speech signal in order to determine the required speech rate. An additional capacity enhancement can be achieved

by introducing network control, which implies that the speech codec has to respond to a network-originated control signal for switching the speech rate to one of a predetermined set of possible rates. The network control procedure for example was addressed by Hanzo *et al.* [371] and Kawashima *et al.* [383]. Specifically, in [371] a novel high-quality, low complexity dual-rate 4.7 kbit/s and 6.5 kbit/s ACELP codec was proposed for indoor communications, which was capable of dropping the associated source rate and speech quality under network control, in order to invoke a more resilient modem mode, amongst less favourable channel conditions. Source-matched binary BCH channel codecs combined with unequal protection diversity- and pilot-assisted 16QAM and 64QAM was employed, in order to accommodate both the 4.7 and the 6.5 kbit/s coded speech bits at a fixed signalling rate of 3.1 kBd. Good communications quality speech was reported in an equivalent speech channel bandwidth of 4 kHz, if the channel Signal-to-Noise Ratio (SNR) and Signal-to-Interference (SIR) of the benign indoors cordless channels were in excess of about 15 and 25 dB for the lower and higher speech quality 16QAM and 64QAM systems, respectively. In [383], Kawashima *et al.* proposed network control procedures for CDMA systems, focusing only on the downlink from the base to the mobile station, where the base station can readily coordinate the coding rate of all users without any significant delay. This network control scheme was based on the so-called M/M/∞/M queueing model applied to a cell under heavy traffic conditions. A modified version of the QCELP codec [379] was used, employing fixed rates of 9.6 kbit/s and 4.8 kbit/s.

Focussing our attention on the associated transmission aspects, in recent years significant research interests have also been devoted to Burst-by-Burst Adaptive Quadrature Amplitude Modulation (BbB-AQAM) transceivers [51]- [391]. The transceiver reconfigures itself on a burst-by-burst basis, depending on the instantaneous perceived wireless channel quality. More explicitly, the associated channel quality of the next transmission burst is estimated and the specific modulation mode, which is expected to achieve the required performance target at the receiver is then selected for the transmission of the current burst. Modulation schemes of different robustness and of different data throughput have also been investigated [392]-[393]. The BbB-AQAM principles have also been applied to Joint Detection Code Division Multiple Access (JD-CDMA) [374, 394] and OFDM [395, 396].

Against the above background, in this section we introduce a novel dual-mode burst-by-burst adaptive speech transceiver scheme, based on the AMR speech codec, Redundant Residue Number System (RRNS) assisted channel coding [397] and Joint Detection aided Code-Division Multiple Access (JD-CDMA) [374]. The mode switching is controlled by the channel quality fluctuations imposed by the time-variant channel, which is not necessarily a desirable scenario. However, we will endeavour to contrive measures in order to mitigate the associated perceptual speech quality fluctuations. The underlying trade-offs associated with employing two speech modes of the AMR standard speech codec in conjunction with a reconfigurable, unequal error protection BPSK/4QAM modem are investigated.

10.5 System Overview

The schematic of the proposed adaptive JD-CDMA speech transceiver is depicted in Figure 10.6. The encoded speech bits generated by the AMR codec at the bit rate of 4.75 or 10.2 kbit/s are first mapped according to their error sensitivities into three protection classes,

Class	RRNS Code	Number of Codewords	databits	Total databits	Total codedbits
4.75kbit/s/BPSK					
I	RRNS$(8,4)$	2	40	95	160
II	RRNS$(8,5)$	1	25		
III	RRNS$(8,6)$	1	30		
10.2kbit/s/4QAM					
I	RRNS$(8,4)$	3	60	205	320
II	RRNS$(8,5)$	1	25		
III	RRNS$(8,6)$	4	120		

Table 10.4: RRNS codes designed for two different modulation modes.

although for simplicity this is not shown explicitly in the figure. The sensitivity-ordered speech bits are then channel encoded using the RRNS encoder [397] and modulated using a re-configurable BPSK or 4QAM based JD-CDMA scheme [51]. We assigned the 4.75 kbit/s speech codec mode to the BPSK modulation mode, while the 10.2 kbit/s speech codec mode to the 4QAM mode. Therefore, this transmission scheme delivers a higher speech quality at 10.2 kbit/s, provided that sufficiently high channel SNRs and SIRs prevail. Furthermore, it can be reconfigured under transceiver control in order to provide an inherently lower, but unimpaired speech quality amongst lower SNR and SIR conditions at the speech rate of 4.75 kbit/s.

Subsequently, the modulated symbols are spread in Figure 10.6 by the CDMA spreading sequence assigned to the user, where a random spreading sequence is used. The Minimum Mean Squared Error Block Decision Feedback Equaliser (MMSE-BDFE) is used as the multiuser detector [374], where perfect Channel Impulse Response (CIR) estimation and perfect decision feedback are assumed. The soft outputs for each user are obtained from the MMSE-BDFE and passed to the RRNS channel decoder. Finally, the decoded bits are mapped back to their original bit protection classes by using a bit-mapper (not shown in Figure 10.6) and the speech decoder reconstructs the original speech information.

In BbB-AQAM/CDMA, in order to determine the best choice of modulation mode in terms of the required trade-off between the BER and throughput, the near instantaneous quality of the channel has to be estimated. The channel quality is estimated at receiver *A* and the chosen modulation mode and its corresponding speech mode are then communicated using explicit signalling to transmitter *B* in a closed-loop scheme, as depicted in Figure 10.6. Specifically, the channel quality estimate is obtained by using the Signal to residual Interference plus Noise Ratio (SINR) metric, which can be calculated at the output of MMSE-BDFE [374].

10.6 Redundant Residue Number System (RRNS) Channel Coding

10.6.1 Overview

In order to improve the performance of the system, we employ the novel family of the so-called Redundant Residue Number System (RRNS) codes for protecting the speech bits, depending on their respective error sensitivities.

Since their introduction, RRNS have been used for constructing fast arithmetics [398, 399]. In this paper, we exploit the error control properties of the non-binary systematic RRNS codes, which - similarly to Reed-Solomon codes - exhibit maximum minimum distance properties [400, 401]. Hence, RRNS codes are similar to Reed Solomon (RS) codes [339]. However, the RRNS codes chosen in our design are more amenable to designing short codes. More explicitly, in the context of RS codes, short codes are derived by inserting dummy symbols into full-length codes. This, however, requires the decoding of the full-length RS-code. By contrast, RRNS codes simply add the required number of redundant symbols. Furthermore, RRNS codes allow us to use the low-complexity technique of residue dropping [401]. Both of these advantages will be augmented during our further discourse.

An RRNS(n, k) code has k so-called residues, which host the original data bits and the additional $(n - k)$ redundant residues can be employed for error correction at the decoder. The coding rate of the code is k/n and the associated error correction capability of the code is $t = \lfloor \frac{n-k}{2} \rfloor$ non-binary residues [400, 401]. At the receiver, both soft decision [397] and residue dropping [402] decoding techniques are employed.

The advantages of the RRNS codes are simply stated here without proof due to lack of space [397, 402]. Since the so-called residues of the RRNS [398, 399] can be computed independently from each other, additional residues can be added at any stage of processing or transmission [403]. This has the advantage that the required coding power can be adjusted according to the prevalent BER of the transmission medium. For example, when the protected speech bits enter the wireless section of the network - where higher BERs prevail than in the fixed network - simply a number of additional redundant residues are computed and concatenated to the message for providing extra protection.

In our design, RRNS codes employing 5 bits per residue have been chosen. Three different RRNS codes having different code rates are used for protecting the three different classes of speech bits. In addition, the RRNS codes employed are also switched in accordance with the modulation modes and speech rates used in our system. In Table 10.4, we have two set of RRNS codes for the BPSK and 4QAM modulation modes. For the most sensitive class I speech bits, we used a RRNS(8,4) code, which has a minimum free distance of $d_{min} = 5$ [397] and a code rate of $1/2$. At the receiver, the soft metric of each received bit was calculated and soft decoding was applied. An extra information residue was added to the RRNS(8,4) code for generating the RRNS(8,5) code for the speech bit protection class II. The extra residue enables us to apply one residue dropping [402], and soft decision decoding. The Class III bits are least protected, using the RRNS(8,6) code, which has a minimum free distance of $d_{min} = 3$ and a code rate of $2/3$. Only soft decision decoding is applied to this code.

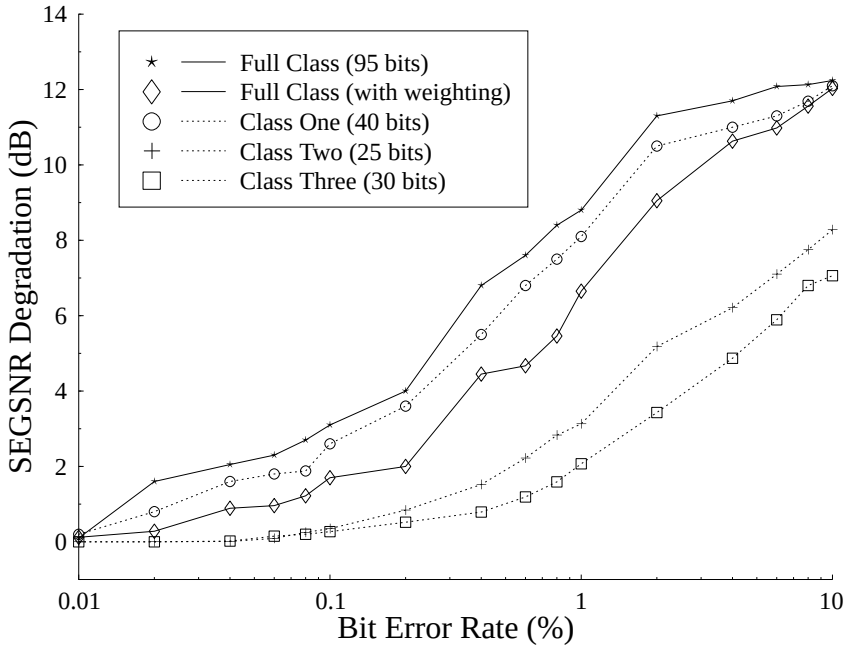

Figure 10.7: SEGSNR degradation versus average BER for the 4.75 kbit/s AMR codec for full-class and triple-class protection systems. When the bits of a specific class were corrupted, bits of the other classes were kept intact.

10.6.2 Source-Matched Error Protection

The error sensitivity of the 4.75 kbit/s AMR codec's source bits was evaluated in Figures 10.3 and 10.5. The same procedures were applied in order to obtain the error sensitivity for the source bits of the 10.2 kbit/s AMR codec. Again, in our system, we employed RRNS channel coding and three protection classes were deemed to constitute a suitable trade-off between a moderate system complexity and high performance. As shown in Table 10.4, three different RRNS codes having different code rates are used for protecting the three different classes of speech bits in a speech frame.

For the 4.75 kbit/s AMR speech codec, we divided the 95 speech bits into three sensitivity classes, Class I, II and III. Class I consists of 40 bits, while Class II and III were allocated 25 and 30 bits, respectively. Then we evaluated the associated SEGSNR degradation inflicted by certain fixed channel BERs maintained in each of the classes using randomly distributed errors, while keeping bits of the other classes intact. The results of the SEGSNR degradations applying random errors are portrayed in Figure 10.7 for both the full-class and the triple-class

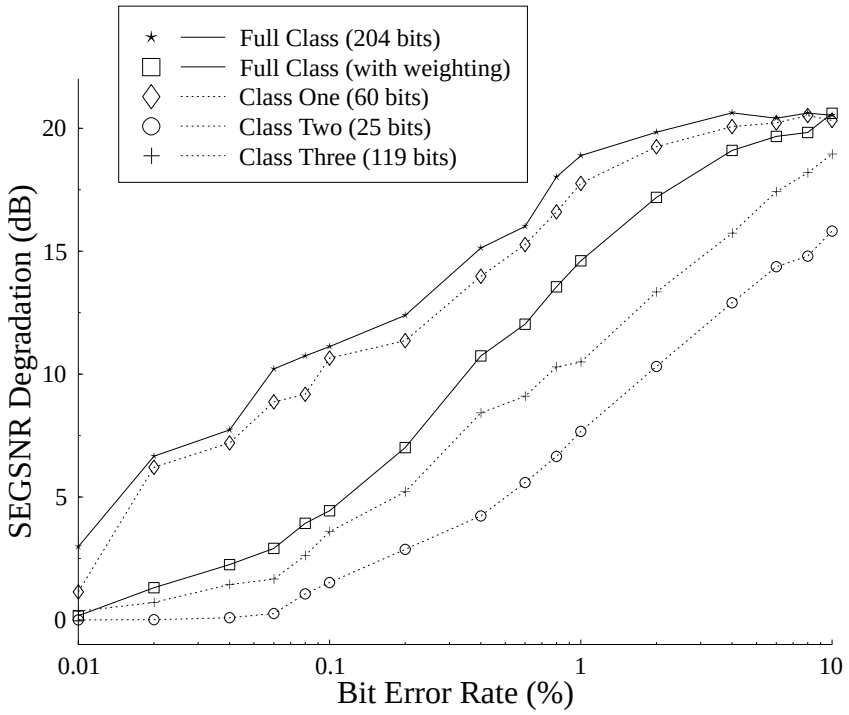

Figure 10.8: SEGSNR degradation versus average BER for the 10.2 kbit/s AMR codec for full class and triple-class protection systems. When the bits of a specific class were corrupted, bits of the other classes were kept intact.

system. It can be seen that Class I, which consists of the 40 most sensitive bits, suffers the highest SEGSNR degradation. Class II and Class III - which are populated mainly with the fixed codebook index bits - are inherently more robust to errors. Note that in the full-class scenario the associated SEGSNR degradation is higher than that of the individual protection classes. This is due to having more errors in the entire 95-bit frame at a fixed BER, compared to the individual protection classes assigned 40, 25 and 30 bits respectively, since upon corrupting a specific class using a fixed BER, the remaining classes remained intact. Hence the BER of the individual protection classes averaged over all the 95 bits was lower, than that of the full-class scenario. For the sake of completeness, we decreased the BER of the full-class scheme so that on average the same number of errors was introduced into the individual classes as well as in the full-class scheme. In this scenario, it can be seen from Figure 10.7 that, as expected, the Class I scheme has the highest SEGSNR degradation, while the sensitivity of the full-class scheme is mediocre.

Similarly, the 204 bits of a speech frame in the 10.2 kbit/s AMR speech codec mode are

divided into three protection classes. Class I is allocated the 60 most sensitive bits, while 25 and 119 bits are distributed to Class II and Class III, in decreasing order of error sensitivity. Their respective SEGSNR degradation results against the BER are presented in Figure 10.8. Due to the fact that the number of bits in Class III is five times higher than in Class II, the error sensitivity of Class III compared to Class II appeared higher. Hence the SEGSNR degradation appears higher for Class III than for Class II, as observed in Figure 10.8. This occurs due to the non-trivial task of finding appropriate channel codes to match the source sensitivities, and as a result, almost 60% of the bits are allocated to Class III. Note that after the RRNS channel coding stage, an additional dummy bit is introduced in Class III, which contains 119 useful speech bits, as shown in Table 10.4. The extra bit can be used as a Cyclic Redundancy Check (CRC) bit for the purpose of error detection. Having considered the source and channel coding aspects, let us now focus our attention on transmission issues.

10.7 Joint Detection Code Division Multiple Access

10.7.1 Overview

Joint detection receivers [404] constitute a class of multiuser receivers that were developed based on conventional channel equalization techniques [51] used for mitigating the effects of Inter-Symbol Interference (ISI). These receivers utilize the Channel Impulse Response (CIR) estimates and the knowledge of the spreading sequences of all the users in order to reduce the level of Multiple Access Interference (MAI) in the received signal.

By concatenating the data symbols of all CDMA users successively, as though they were transmitted by one user, we can apply the principles of conventional single-user channel equalization [51] to multiuser detection. In our investigations, we have used the MMSE-BDFE proposed by Klein *et al.* [404], where the multiuser receiver aims to minimize the mean square error between the data estimates and the transmitted data. A feedback process is incorporated, where the previous data estimates are fed back into the receiver in order to remove the residual interference and to assist in improving the BER performance.

10.7.2 Joint Detection Based Adaptive Code Division Multiple Access

In QAM [51], n bits are grouped to form a signalling symbol and $m = 2^n$ different symbols convey all combinations of the n bits. These m symbols are arranged in a modulation constellation to form the m-QAM scheme. In the proposed system we used the BbB-AQAM/CDMA modes of BPSK (2-QAM) and 4QAM, conveying 1 and 2 bits per symbol, respectively. However, for a given channel SNR, the BER performance degrades upon switching from BPSK to 4QAM, whilst doubling the throughput.

Previous research in BbB-AQAM schemes designed for TDMA transmissions has been carried out by Webb and Steele [405]; Sampei, Komaki and Morinaga [391]; Goldsmith and Chua [406]; as well as Torrance *et al.* [407]. This work has been extended to wideband channels, where the received signal also suffers from ISI in addition to amplitude and phase distortions due to the fading channel. The received signal strength is not a good indicator of the wideband channel's instantaneous quality, since the signal is also contaminated by ISI and co-channel interference. Wong *et al.* [408] proposed a wideband BbB-AQAM scheme, where a channel equalizer was used for mitigating the effects of ISI on the CIR estimate.

Parameter	Value
Channel type	COST 207 Bad Urban (BU)
Paths in channel	7
Doppler frequency	80 Hz
Spreading factor	16
Chip rate	2.167 MBaud
JD block size	26 symbols
Receiver type	MMSE-BDFE
AQAM type	Dual-mode (BPSK, 4QAM)
Channel codec	Triple-class RRNS
Channel-coded Rate	8/16 kbit/s
Speech Codec	AMR (ACELP)
Speech Rate	4.75/10.2 kbit/s
Speech Frame Length	20 ms

Table 10.5: Transceiver Parameters

Here we propose to combine joint detection CDMA [404] with AQAM, by modifying the approach used by Wong *et al.* [408]. Joint detection is particularly suitable for combining with AQAM, since the implementation of the joint detection algorithms does not require any knowledge of the modulation mode used [374]. Hence the associated complexity is independent of the modulation mode used.

In order to choose the most appropriate BbB-AQAM/CDMA mode for transmission, the SINR at the output of the MMSE-BDFE was estimated by modifying the SINR expression given in [404] exploiting the knowledge of the transmitted signal amplitude, g, the spreading sequence and the CIR. The data bits and noise values were assumed to be uncorrelated. The average output SINR was calculated for each transmission burst of each user. The conditions used for switching between the two AQAM/JD-CDMA modes were set according to their target BER requirements as:

$$\text{Mode} = \begin{cases} \text{BPSK} & \text{SINR} < t_1 \\ \text{4QAM} & t_1 \leq \text{SINR} \end{cases}, \tag{10.1}$$

where t_1 represents the switching threshold between the two modes.

With the system elements described, we now focus our attention on the overall performance of the adaptive transceiver proposed.

10.8 System Performance

The simulation parameters used in our AQAM/JD-CDMA system are listed in Table 10.5. The channel profile used was the COST 207 Bad Urban (BU) channel [409] consisting of seven paths, where each path was faded independently at a Doppler frequency of 80 Hz.

The BER performance of the proposed system is presented in Figures 10.9, 10.10 and 10.11. Specifically, Figure 10.9 portrays the BER performance using the 4QAM modulation

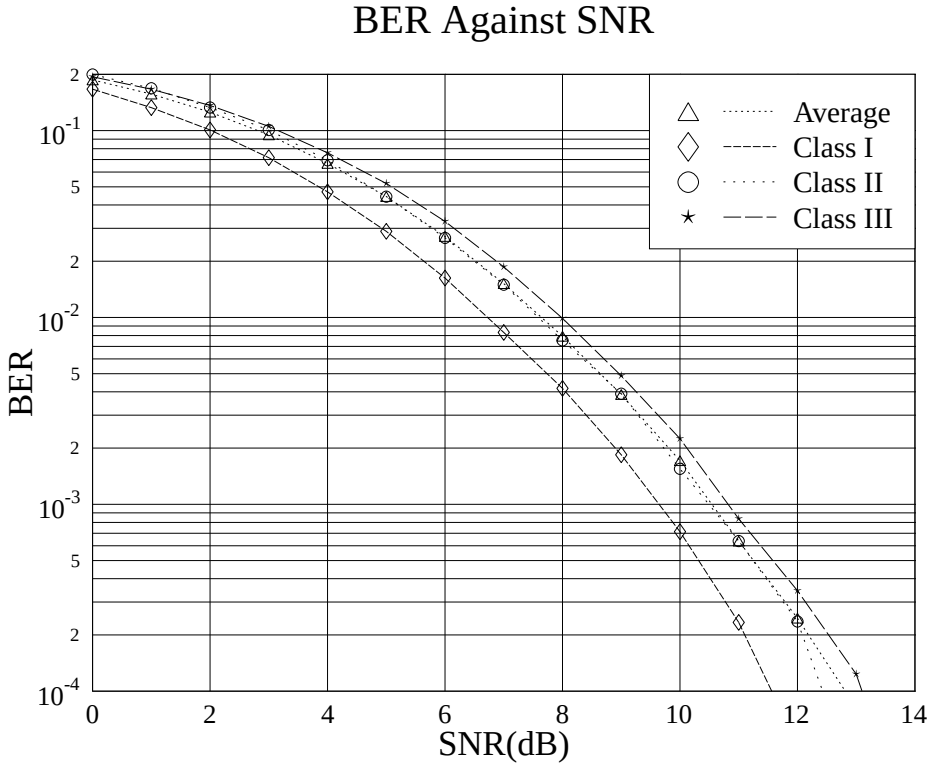

Figure 10.9: BER performance of 4QAM/JD-CDMA over the COST 207 BU channel of Table 10.5 using the RRNS codes of Table 10.4.

mode and the RRNS codes of Table 10.4 for a two-user JD-CDMA speech transceiver. As seen in Table 10.4 of Section 10.6.2, three different RRNS codes having different code rates are used for protecting the three different classes of speech bits in the speech codec. The BER of the three protection classes is shown together with the average BER of the channel coded bits versus the channel SNR. The number of bits in these protection classes was 60, 25 and 120, respectively. As expected, the Class I subchannel exhibits the highest BER performance, followed by the Class II and Class III subchannels in decreasing order of BER performance. The corresponding BER results for the BPSK/JD-CDMA mode are shown in Figure 10.10.

In Figure 10.11, the average BER performance of the coded fixed-mode BPSK/JD-CDMA and 4QAM/JD-CDMA systems is presented along with that of the twin-mode AQAM/JD-CDMA system supporting two users and assuming zero-latency modem mode signalling. The performance of the AQAM scheme was evaluated by analyzing the BER and the throughput expressed in terms of the average number of Bits Per Symbol (BPS) transmitted. The

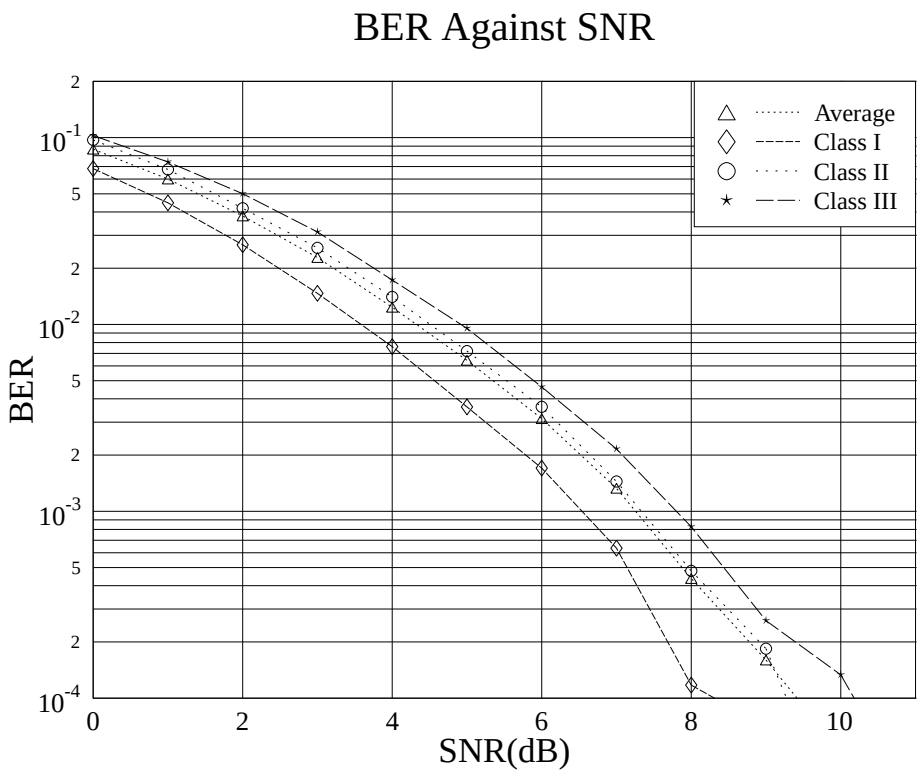

Figure 10.10: BER performance of BPSK/JD-CDMA over the COST 207 BU channel of Table 10.5 using the RRNS codes of Table 10.4.

BER curve has to be read by referring to the vertical-axis at the left of the figure, while the BPS throughput curve is interpreted by referring to the vertical-axis at the right that is labelled BPS. At low channel SNRs the BER of the AQAM/JD-CDMA scheme mirrored that of BPSK/JD-CDMA, which can be explained using Figure 10.12. In Figure 10.12, the Probability Density Functions (PDF) of the AQAM/JD-CDMA modes versus channel SNR are plotted. As mentioned earlier, the results were obtained using a switching threshold of 10.5 dB. We can see from the figure that at low average channel SNRs (< 6 dB), the mode switching threshold of 10.5 dB instantaneous SNR was seldom reached, and therefore BPSK/JD-CDMA was the predominant mode. Hence, the performance of the AQAM/JD-CDMA scheme was similar to BPSK/JD-CDMA. However, as the channel SNR increased, the BER performance of AQAM/JD-CDMA became better than that of BPSK/JD-CDMA, as shown in Figure 10.11. This is because the 4QAM mode is employed more often, reducing the probability of using BPSK, as shown in Figure 10.12. Since the mean BER of the system

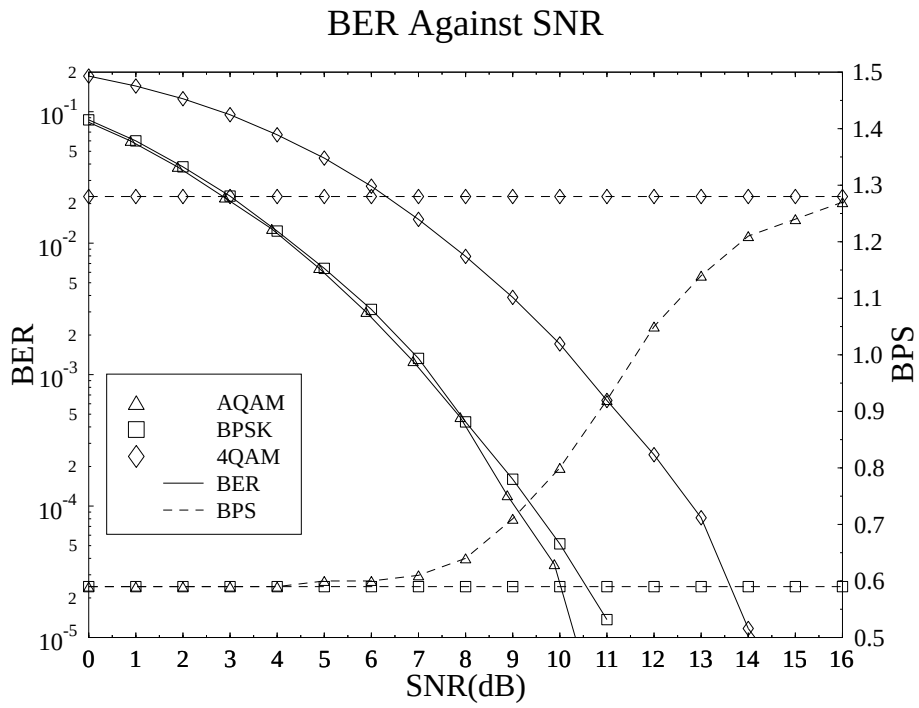

Figure 10.11: BER and BPS comparisons for fixed mode BPSK and 4QAM as well as for the AQAM/JD-CDMA system, using the RRNS codes of Table 10.4. The switching threshold for AQAM was set to 10.5 dB and the simulation parameters are listed in Table 10.5.

is the ratio of the total number of bit errors to the total number of bits transmitted, the mean BER will decrease with a decreasing number of bit errors or with an increasing number of transmitted bits. For a fixed number of symbols transmitted, the total number of transmitted bits in a frame is constant for fixed mode BPSK/JD-CDMA, while for AQAM/JD-CDMA the total number of transmitted bits increased, when the 4QAM/JD-CDMA mode was used. Consequently, the average BER of the AQAM/JD-CDMA system was lower than that of the fixed-mode BPSK/JD-CDMA scheme.

The BPS throughput performance curve is also plotted in Figure 10.11. As expected, the number of BPS of both BPSK and 4QAM is constant for all channel SNR values. They are limited by the modulation scheme used and the coding rate of the RRNS codes seen in Table 10.4. For example, for 4QAM we have 2 BPS, but the associated channel code rate is $205/320$, as shown in Table 10.4, hence the effective throughput of the system is $2 \times \frac{205}{320} = 1.28$ BPS. For AQAM/JD-CDMA, we can see from Figure 10.11 that the throughput is similar to that of BPSK/JD-CDMA at low channel SNRs. However, as the average

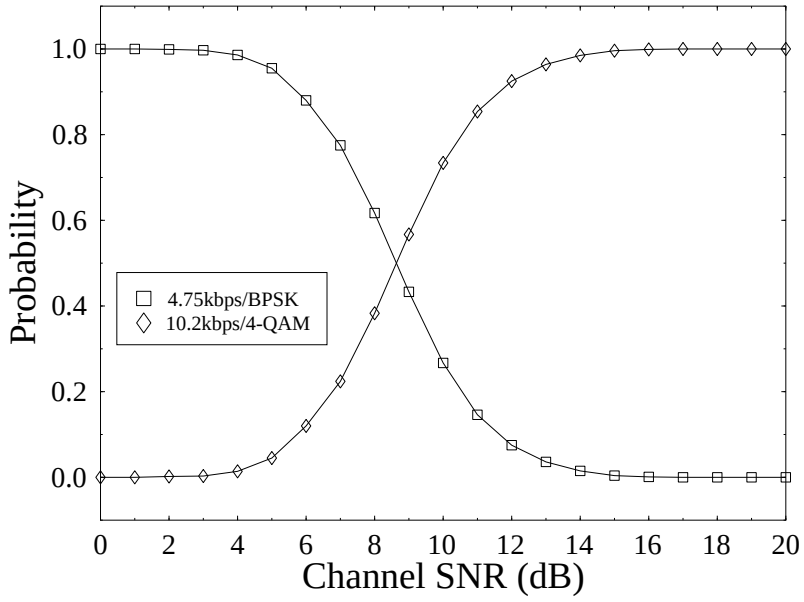

Figure 10.12: The probability of each modulation mode being chosen for transmission in a twin-mode (BPSK, 4QAM), two-user AQAM/JD-CDMA system using the parameters of Table 10.5.

channel SNR increased, more and more frames were transmitted using 4QAM/JD-CDMA and the average throughput increased gradually. At high average SNRs, the throughput of AQAM/JD-CDMA became similar to that of the 4QAM/JD-CDMA scheme.

The overall SEGSNR versus channel SNR performance of the proposed speech transceiver is displayed in Figure 10.13. Observe that the source sensitivity-matched triple-class 4.75 kbit/s BPSK/JD-CDMA system requires a channel SNR in excess of about 8 dB for nearly unimpaired speech quality over the COST207 BU channel of Table 10.5. When the channel SNR was in excess of about 12 dB, the 10.2 kbit/s 4QAM/JD-CDMA system outperformed the 4.75 kbit/s BPSK/JD-CDMA scheme in terms of both objective and subjective speech quality. Furthermore, at channel SNRs around 10 dB, where the BPSK and 4QAM SEGSNR curves cross each other in Figure 10.13, it was preferable to use the inherently lower quality but unimpaired mode of operation. In the light of these findings, the application of the AMR speech codec in conjunction with AQAM constitutes an attractive trade-off in terms of providing users with the best possible speech quality under arbitrary channel conditions. Specifically, the 10.2 kbit/s 4QAM/JD-CDMA scheme has a higher source bit rate and thus exhibits a higher SEGSNR under error-free conditions. The 4.75 kbit/s BPSK/JD-CDMA scheme exhibits a lower source bit rate and correspondingly lower speech quality under error-

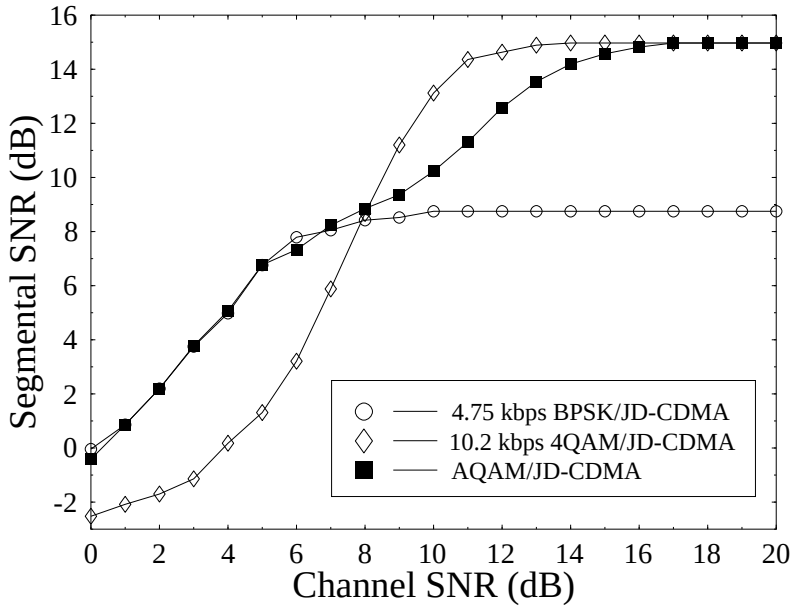

Figure 10.13: SEGSNR versus channel SNR

free conditions. However, due to its less robust 4QAM modulation mode, the 10.2 kbit/s 4QAM/JD-CDMA scheme is sensitive to channel errors and breaks down under hostile channel conditions, where the 4.75 kbit/s BPSK/JD-CDMA scheme still exhibits robust operation, as illustrated in Figure 10.13.

In the context of Figure 10.13 ideally a system is sought that achieves a SEGSNR performance, which follows the envelope of the SEGSNR curves of the individual BPSK/JD-CDMA and 4QAM/JD-CDMA modes. The SEGSNR performance of the AQAM system is also displayed in Figure 10.13. We observe that AQAM provides a smooth evolution across the range of channel SNRs. At high channel SNRs, in excess of 12-14 dB, the system operates predominantly in the 4QAM/JD-CDMA mode. As the channel SNR degrades below 12 dB, some of the speech frames are transmitted in the BPSK/JD-CDMA mode, which implies that the lower quality speech rate of 4.75 kbit/s is employed. This results in a slightly degraded average speech quality, while still offering a substantial SEGSNR gain compared to the fixed-mode 4.75 kbit/s BPSK/JD-CDMA scheme. At channel SNRs below 10 dB, the performance of the 10.2 kbit/s 4QAM/JD-CDMA mode deteriorates due to the occurrence of a high number of channel errors, inflicting severe SEGSNR degradations. In these hostile conditions, the 4.75 kbit/s BPSK/JD-CDMA mode provides a more robust performance associated with a better speech quality. With the advent of the AQAM/JD-CDMA mode switching

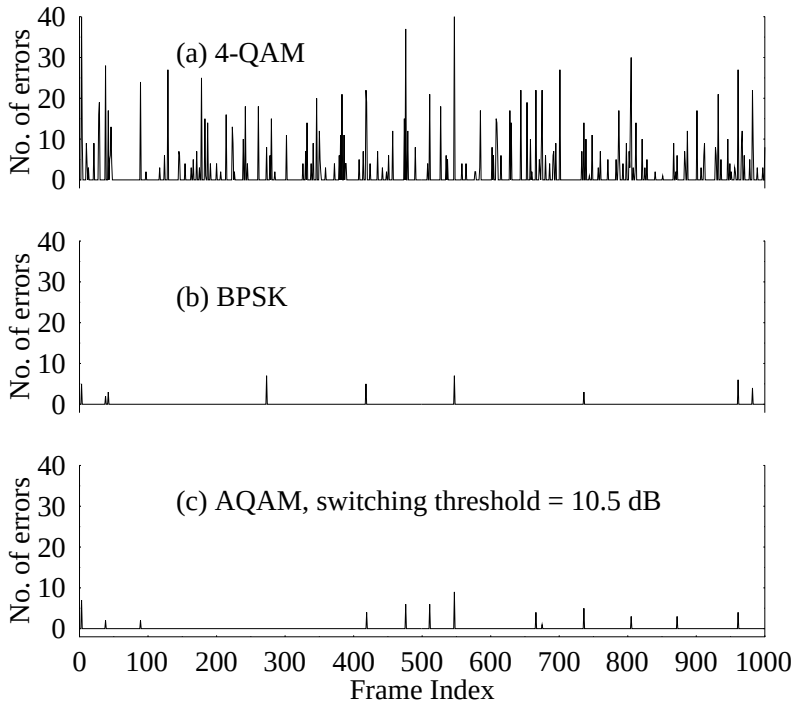

Figure 10.14: The comparison of the number of errors per frame versus 20ms frame index for the (a) 4QAM, (b) BPSK and (c) AQAM/JD-CDMA systems with a switching threshold of 10.5 dB, at channel SNR = 10 dB for 1000 frames over the COST207 BU channel of Table 10.5.

regime the transceiver exhibits a less bursty error distribution, than that of the conventional fixed-mode 4QAM modem, as it can be seen in Figure 10.14, where the error events of the BPSK/JD-CDMA scheme are also displayed.

The benefits of the proposed dual-mode transceiver are further demonstrated by Figure 10.15, consisting of three graphs plotted against the speech frame index, giving an insightful characterisation of the adaptive speech transceiver. Figure 10.15(a) shows a speech segment of 30 frames. In the AMR codec, a speech frame corresponds to a duration of 20 ms. In Figure 10.15(b), the SEGSNR versus frame index performance curves of the BPSK, 4QAM and AQAM/JD-CDMA schemes are shown, in both error-free and channel-impaired scenarios. The SINR at the output of the MMSE-BDFE is displayed in Figure 10.15(c). The adaptation of the modulation mode is also shown in Figure 10.15(c), where the transceiver switches to the BPSK or 4QAM mode according to the estimated SINR using the switching threshold set to 10.5 dB.

When transmitting in the less robust 4QAM mode using the higher-rate speech mode of 10.2 kbit/s, a sudden steep drop in the channel conditions - as portrayed at Frame 1 in Figure 10.15 - results in a high number of transmission errors, as also illustrated in Figure 10.14(a).

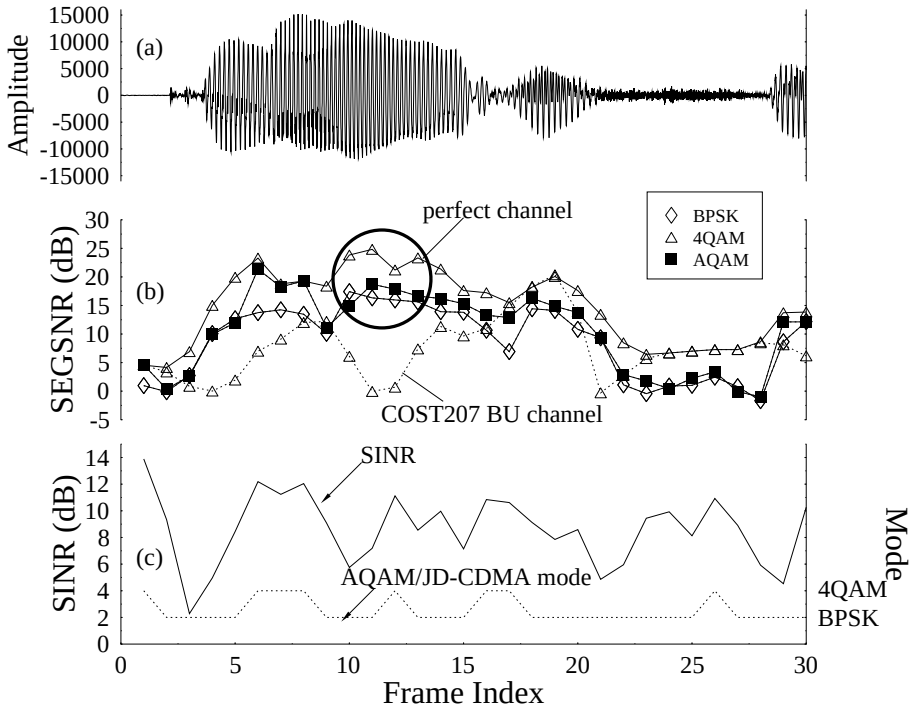

Figure 10.15: Characteristic waveforms of the adaptive system. (a) Time-domain speech signal for frame indices between 0 and 30; (b) SEGSNR in various transceiver modes; (c) SINR versus time and transceiver modes versus time over the COST207 BU channel of Table 10.5.

This happens to occur during the period of voice onset in Figure 10.15, resulting in the corruption of the speech frame, which has the effect of inflicting impairments to subsequent frames due to the error propagation effects of various speech bits, as alluded to in Section 10.3. It can be seen in Figure 10.15 that the high number of errors inflicted in the 4QAM mode during voiced speech segments caused a severe SEGSNR degradation at frame index 10 and the 10.2 kbit/s speech codec never fully recovered, until the channel conditions expressed in terms of the SINR in Figure 10.15(c) improved. On the other hand, the significantly more robust 4.75 kbit/s BPSK/JD-CDMA scheme performed well under these hostile channel conditions, encountering a low number of errors in Figure 10.14(b), while transmitting at a lower speech rate, hence at an inherently lower speech quality. For the sake of visual clarity, the performance curves of BPSK/JD-CDMA and AQAM/JD-CDMA were not displayed in Figure 10.15(b) for the channel-impaired scenarios, since their respective graphs are almost identical to that of the error-free speech SEGSNR curves.

The benefits of the proposed dual-mode transceiver are also demonstrated by Figure 10.16, which shares the same graphs arrangement as described earlier for Figure 10.15 but at a different frame index range between 300 and 330. It can be seen in Figure 10.16 that a

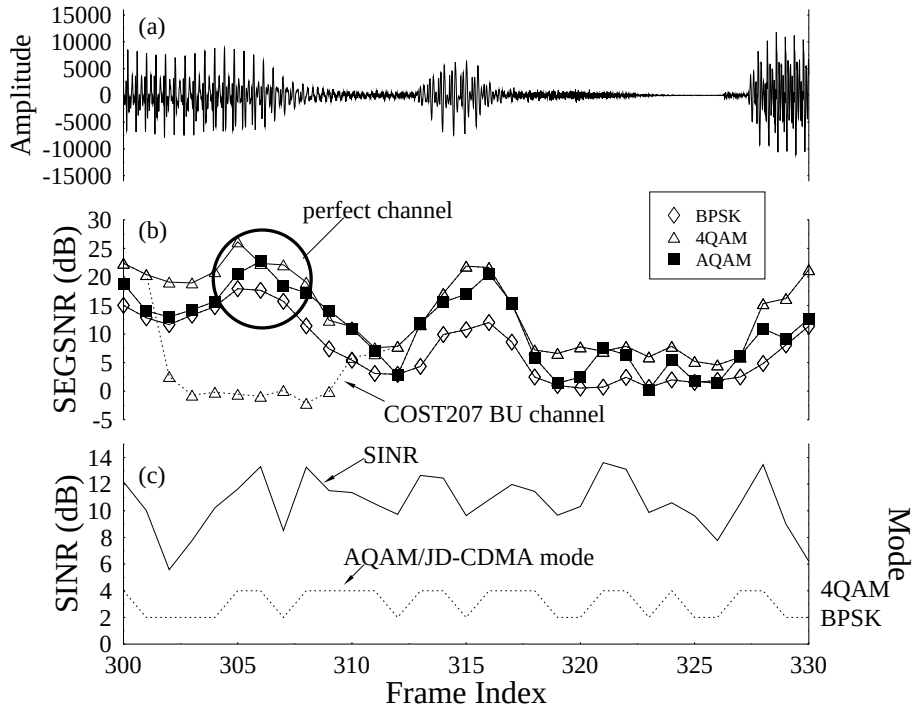

Figure 10.16: Characteristic waveforms of the adaptive system. (a) Time-domain speech signal for frame indices between 300 and 330; (b) SEGSNR in various transceiver modes; c) SINR versus time and transceiver modes versus time over the COST207 BU channel of Table 10.5.

sudden steep drop in the channel conditions at frame index 300 during the 4QAM mode, caused a severe SEGSNR degradation for the voiced speech segments and the 10.2 kbit/s speech codec never recovered until the channel conditions improved.

10.8.1 Subjective Testing

Informal listening tests were conducted, in order to assess the performance of the AQAM/JD-CDMA scheme in comparison to the fixed-mode BPSK/JD-CDMA and 4QAM/JD-CDMA schemes. It is particularly revealing to investigate, how the AQAM/JD-CDMA scheme performs in the intermediate channel SNR region between 7 dB and 11 dB. The speech quality was assessed using pairwise comparison tests. The listeners were asked to express a preference between two speech files A or B or neither. A total of 12 listeners were used in the pairwise comparison tests. Four different utterances were employed during the listening tests, where the utterances were due to a mixture of male and female speakers having American accents. Table 10.6 details some of the results of the listening tests.

Through the listening tests we found that for the fixed-mode BPSK/JD-CDMA scheme unimpaired perceptual speech quality was achieved for channel SNRs in excess of 7 dB.

Speech Material A	Speech Material B	Preference		
		A (%)	B (%)	Neither (%)
4.75 kbit/s (Error free)	10.2 kbit/s (Error free)	4.15	66.65	29.2
AQAM(9dB)	4QAM(9dB)	100	0.00	0.00
AQAM(9dB)	4QAM(11dB)	8.3	50.0	41.7
AQAM(9dB)	BPSK(9dB)	37.5	16.65	45.85
AQAM(12dB)	4QAM(12dB)	4.15	20.85	75.0
AQAM(12dB)	4QAM(13dB)	8.3	25.0	66.7
AQAM(12dB)	BPSK(12dB)	41.65	8.3	50.05

Table 10.6: Details of the listening tests conducted using the pairwise comparison method, where the listeners were given a choice of preference between two speech files coded in different transmission scenarios.

With reference to Figure 10.13, when the channel conditions degraded below 7 dB, the speech quality became objectionable due to the preponderance of channel errors. For the fixed mode 4QAM/JD-CDMA scheme, the channel SNR threshold was 11 dB, below which the speech quality started to degrade. The perceptual performance of AQAM/JD-CDMA was found superior to that of 4QAM/JD-CDMA at channel SNRs below 11 dB. Specifically, it can be observed from Table 10.6 that all the listeners preferred the AQAM/JD-CDMA scheme at a channel SNR of 9 dB due to the associated high concentration of channel errors in the less robust 4QAM/JD-CDMA scheme at the same channel SNR, resulting in a perceptually degraded reconstructed speech quality.

More explicitly, we opted for investigating the AQAM/JD-CDMA scheme at a channel SNR of 9 dB, since - as shown in Figure 10.12 - this SNR value falls in the transitory region between BPSK/JD-CDMA and 4QAM/JD-CDMA. As the channel conditions improved to an SNR in excess of 11 dB, the 4QAM/JD-CDMA scheme performed slightly better, than AQAM/JD-CDMA due to its inherently higher SEGSNR performance under error free conditions. Nonetheless, the AQAM/JD-CDMA scheme provided a good perceptual performance, as exemplified in Table 10.6 at a channel SNR of 12 dB, in comparison to the 4QAM/JD-CDMA scheme at the channel SNRs of both 12 dB and 13 dB. Here, only about twenty percent of the listeners preferred the 4QAM/JD-CDMA scheme to the AQAM/JD-CDMA scheme, while the rest suggested that both sounded very similar. It can also be observed from Table 10.6 that the AQAM/JD-CDMA scheme performed better than BPSK/JD-CDMA for a channel SNR of 7 dB and above, while in the region below 7 dB, AQAM/JD-CDMA has a similar perceptual performance to that of BPSK/JD-CDMA. As shown in Table 10.7, we found that changing the mode switching frequency for every 1, 10 or 100 frames does not impair the speech quality either in objective SEGSNR terms or in terms of informal listening tests.

10.9 A Turbo-Detected Unequal Error Protection Irregular Convolutional Coded AMR Transceiver

J. Wang, N. S. Othman, J. Kliewer, L. L. Yang and L. Hanzo

Frame Switching Frequency	SEGSNR (dB)
1	11.38
10	11.66
100	11.68

Table 10.7: Frame switching frequency versus SEGSNR

10.9.1 Motivation

Since the different bits of multimedia information, such as speech and video, have different error sensitivity, efficient unequal-protection channel coding schemes have to be used to ensure that the perceptually more important bits benefit from more powerful protection. Furthermore, in the context of turbo detection the channel codes should also match the characteristics of the channel for the sake of attaining a good convergence performance. In this section we address this design dilemma by using irregular convolutional codes (IRCCs) which constitute a family of different-rate subcodes. we benefit from the high design flexibility of IRCCs and hence excellent convergence properties are maintained while having unequal error protection capabilities matched to the requirements of the source. An EXIT chart based design procedure is proposed and used in the context of protecting the different-sensitivity speech bits of the wideband AMR speech codec. As a benefit, the unequal-protection system using IRCCs exhibits an SNR advantage of about 0.4 dB over the equal-protection system employing regular convolutional codes, when communicating over a Gaussian channel. We will also demonstrate that irregular Convolutional codes exhibit excellent convergence properties in the context of iterative decoding, whilst having an unequal error protection capability, which is exploited in this contribution to protect the different-sensitivity speech bits of the wideband AMR speech codec. As a benefit, the unequal-protection system exhibits an SNR advantage of about 0.3 dB over the equal-protection system, when communicating over a Guassian channel.

Source encoded information sources, such as speech, audio or video, typically exhibit a non-uniform error sensitivity, where the effect of a channel error may significantly vary from one bit to another [410, 411]. Hence unequal error protection (UEP) is applied to ensure that the perceptually more important bits benefit from more powerful protection. In [340], the speech bits were protected by a family of Rate-Compatible Punctured Convolutional (RCPC) codes [341] whose error protection capabilities had been matched to the bit-sensitivity of the speech codec. Different-rate RCPC codes were obtained by puncturing the same mother code, while satisfying the rate-compatibility restriction. However, they were not designed in the context of turbo detection. Other schemes using a serially concatenated system and turbo processing were proposed in [342, 343], where the UEP was provided by two different-rate convolutional codes.

Recently, Tüchler *et al.* [344, 345] studied the construction of irregular convolutional codes (IRCCs) and proposed several design criteria. These IRCCs consisted of a family of convolutional codes having different code rates and were specifically designed with the aid of extrinsic information transfer (EXIT) charts [346] invoked, for the sake of improving the convergence behaviour of iteratively decoded serially concatenated systems. In general, EXIT chart analysis assumes having a long interleaver block lengths. However, it was shown in [345]

that by using an appropriate optimization criterion, the concatenated system is capable of performing well even for short interleaver block lengths. Since the constituent codes have different coding rates, the resultant IRCC is capable of providing unequal error protection (UEP).

A novel element of this section is that UEP and EXIT chart based code optimization will be jointly carried out and successfully applied for improving the achievable robustness of speech transmission. We propose a serially concatenated turbo transceiver using an IRCC as the outer code for the transmission of Adaptive Multi-Rate Wideband (AMR-WB) coded speech. Rather than being decoded separately, the constituent codes of the IRCC are decoded jointly and iteratively by exchanging extrinsic information with the inner code. The IRCC is optimized to match the characteristics of both the speech source codec and those of the channel, so that UEP is achieved while maximizing the iteration gain attained.

In contrast to the error sensitivity of the narrow-band AMR codec characterized in Section 10.3, that of the AMR-WB speech codec will characterized In Section 10.9.2, while our system model will be introduced in Section 10.9.3, followed by Section 10.9.4, which describes the design procedure of IRCCs. An IRCC design example is provided in Section 10.9.5. Our performance results are presented in Section 10.9.6, while Section **??** concludes the section.

10.9.2 The AMR-WB Codec's Error Sensitivity

The AMR-WB speech codec is capable of supporting bit rates varying from 6.6 to 23.85 kbit/s and it has become a 3GPP and ITU-T standard, which provides a superior speech quality in comparison to the conventional telephone-bandwith voice codecs [347]. Each AMR-WB frame represents 20 ms of speech, producing 317 bits at a bitrate of 15.85 kbps plus 23 bits of header information per frame. The codec parameters in each frame include the so-called imittance spectrum pairs (ISPs), the adaptive codebook delay (pitch delay), the algebraic codebook excitation index and the jointly vector quantized pitch gains as well as algebraic codebook gains.

Most source coded bitstreams contain certain bits that are more sensitive to transmission errors than others. A common approach for quantifying the sensitivity of a given bit is to consistently invert this bit in every speech frame and evaluate the associated Segmental SNR (SegSNR) degration [410]. The error sensitivity of the various encoded bits in the AMR-WB codec determined in this way is shown in Fig. 10.17. The results are based on speech samples taken from the EBU SQAM (Sound Quality Assessment Material) CD, sampled at 16 kHz and encoded at 15.85 kbps. It can be observed that the bits representing the ISPs, the adaptive codebook delay, the algebraic codebook index and the vector quantized gain are fairly error sensitive. By contrast, the least sensitive bits are related to the fixed codebook's excitation pulse positions. Statistically, about 10% (35/340) of the bits in a speech frame will cause a SegSNR degration in excess of 10 dB, and about 8% (28/340) of the bits will inflict a degration between 5 and 10 dB. Furthermore, the error-free reception of the 7% (23/340) header information is in general crucial for the adequate detection of speech.

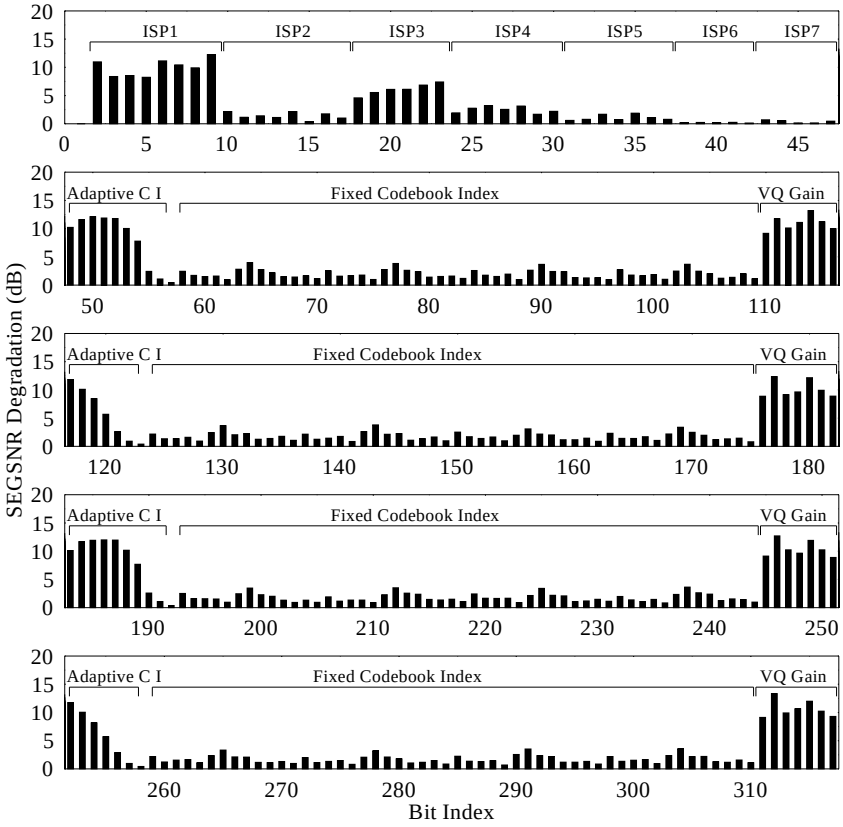

Figure 10.17: SegSNR degrations versus bit index due to inflicting 100% BER in the 317-bit, 20 ms AMR-WB frame

10.9.3 System Model

Fig. 10.18 shows the system's schematic diagram. At the transmitter, each of the K-bit speech frame is protected by a serially concatenated channel code consisting of an outer code (Encoder I) and an inner code (Encoder II) before transmission over the channel, resulting in an overall coding rate of R. At the receiver, iterative decoding is performed with advent of extrinsic information exchange between the inner code (Decoder II) and the outer code (Decoder I). Both decoders employ the a-*posteriori* probability (APP) decoding algorithm, e.g., the BCJR algorithm [348]. After F number of iterations, the speech decoder is invoked in order to reconstruct the speech frame.

According to the design rules of [349], the inner code of a serially concatenated system should be recursive to enable interleaver gain. Furthermore, it has been shown in [350] that for binary erasure channels (BECs) and block lengths tending to infinity the inner code should have rate-1 to achieve capacity. Experiments have shown that this approximately holds also for AWGN channels [344,345]. For the sake of simplicity, we opted for employing a memory-1 recursive convolutional code having a generator polynomial of $1/(1+D)$, which is actually

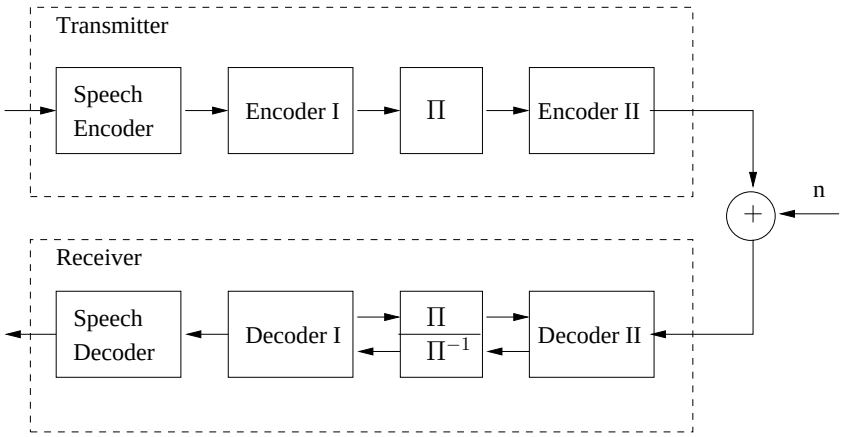

Figure 10.18: System Model

a simple accumulator. Hence the decoding complexity of the inner code is extremely low. In the proposed system, we use an IRCC as the outer code, while in the benchmarker system, we use a regular non-systematic convolutional (NSC) code as the outer code. BPSK modulation and encountering an AWGN channel are assumed.

10.9.4 Design of Irregular Convolutional Codes

An IRCC is constructed from a family of P subcodes. First, a rate-r convolutional mother code C_1 is selected and the $(P-1)$ other subcodes C_k of rate $r_k > r$ are obtained by puncturing. Let L denote the total number of encoded bits generated from the K input information bits. Each subcode encodes a fraction of $\alpha_k r_k L$ information bits and generates $\alpha_k L$ encoded bits. Given the target code rate of $R \in [0,1]$, the weighting coefficient α_k has to satisfy:

$$1 = \sum_{k=1}^{P} \alpha_k, \; R = \sum_{k=1}^{P} \alpha_k r_k, \text{ and } \alpha_k \in [0,1], \; \forall k. \tag{10.2}$$

For example, in [345] a family of $P = 17$ subcodes were constructed from a systematic, rate-1/2, memory-4 mother code defined by the generator polynomial $(1, g_1/g_0)$, where $g_0 = 1 + D + D^4$ is the feedback polynomial and $g_1 = 1 + D^2 + D^3 + D^4$ is the feedforward one. Higher code rates may be obtained by puncturing, while lower rates are created by adding more generators and by puncturing under the contraint of maximizing the achievable free distance. The two additional generators used are $g_2 = 1 + D + D^2 + D^4$ and $g_3 = 1 + D + D^3 + D^4$. The resultant 17 subcodes have coding rates spanning from $0.1, 0.15, 0.2, \cdots$, to 0.9.

The IRCC constructed has the advantage that the decoding of all subcodes may be performed using the same mother code trellis, except that at the beginning of each block of $\alpha_k r_k L$ trellis sections corresponding to the subcode C_k, the puncturing pattern has to be restarted. Trellis termination is necessary only after all of the K information bits have been

encoded.

We now optimize the iterative receiver by means of EXIT charts [346], which is capable of predicting the performance of an iterative receiver by examining the extrinsic information transfer function of each of the component devices independently.

For the outer decoder (Decoder I), denote the mutual information between the a *priori* input A and the transmitted code bits C as $I_{A1} = I(C;A)$, while the mutual information between the extrinsic output E and the transmitted code bits C is denoted as $I_{E1} = I(C;E)$. Then the transfer function of Decoder I can be defined as:

$$I_{E1} = T_I(I_{A1}), \tag{10.3}$$

which maps the input variable I_{A1} to the output variable I_{E1}. Similarly, for the inner decoder (Decoder II), we denote the mutual information between the a *priori* input A and the transmitted information bits X as $I_{A2} = I(X;A)$. Furthermore, we denote the mutual information between the extrinsic output E and the transmitted information bits X as $I_{E2} = I(X;E)$. Note that the extrinsic output of the inner code also depends on the channel SNR or E_b/N_0. Hence the transfer function of the inner code is defined as

$$I_{E2} = T_{II}(I_{A2}, E_b/N_0). \tag{10.4}$$

The transfer functions can be obtained by using the histogram-based LLR measurements as proposed in [346] or the simplified method as proposed in [351].

When using IRCCs, the transfer function of an IRCC can be obtained from those of its subcodes. Denote the transfer function of the subcode k as $T_{I,k}(i)$. Assuming that the trellis fractions of the subcodes do not significantly interfere with each other, which might change the associated transfer characteristics, the transfer function $T_I(i)$ of the target IRCC is the weighted superposition of the transfer function $T_{I,k}(i)$ [345], yielding,

$$T_I(i) = \sum_{k=1}^{P} \alpha_k T_{I,k}(i). \tag{10.5}$$

Note that in iterative decoding, the extrinsic output $E2$ of Decoder II becomes the a *priori* input $A1$ of Decoder I and vice versa. Given the transfer function, $T_{II}(i, E_b/N_0)$, of the inner code, and that of the outer code $T_I(i)$, the extrinsic information I_{E1} at the output of Decoder I after the ith iteration can be calculated using the recursion of:

$$\mu_i = T_I(T_{II}(\mu_{i-1}, E_b/N_0)), \;\; i = 1, 2, \ldots, \tag{10.6}$$

with $\mu_0 = 0$, i.e., assuming the absence of a *priori* input for Decoder II at the commencement of iterations.

Generally, interactive speech communication systems require a low delay, and hence a short interleaver block length. And the number of iterations for the iterative decoder is also limited due to the constraint of complexity. It has been found [345] that EXIT charts may provide a reasonable convergence prediction for the first couple of iterations even in the case of short block lengths. Hence, we fixed the transfer function of the inner code for a given E_b/N_0 value yielding $T_{II}(i) = T_{II}(i, E_b/N_0)$, and optimized the weighting coefficients $\{\alpha_k\}$ of the outer IRCC for the sake of obtaining a transfer function $T_I(i)$ that specifically maximizes the

extrinsic output after exactly F number of iterations [345], which is formulated as:

$$\text{maximize } \mu_i = T_I(T_{II}(\mu_{i-1})), \quad i = 1, 2, \ldots, F, \tag{10.7}$$

with $\mu_0 = 0$.

Additionally, considering the non-uniform error sensitivity of the speech source bits characterized in Figurefig:amr-sensitivity, we may intentionally enhance the protection of the more sensitive source data bits by using strong subcodes, thus imposing the source constraints of:

$$\sum_{k=k_1}^{k_2} \alpha_k r_k / R \geq x\%, \quad 1 \leq k_1 \leq k_2 \leq P, \quad 0 \leq x \leq 100, \tag{10.8}$$

which implies that the percentage of the speech source bits protected by the subcodes k_1 to k_2 is at least $x\%$.

Finally, our task is to find a weight vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_P]^T$, so that eq. (10.7) is maximized, while satisfying the constraints of eq. (10.2) and eq. (10.8). This optimization problem can be solved by slightly modifying the procedure proposed in [345], as it will be illustrated by the following example.

10.9.5 An Example Irregular Convolutional Code

We assume the overall system coding rate to be $R = 0.5$. As stated in Section 10.9.3, the inner code has a unitary code rate, hence all the redundancy is assigned to the outer code. We use a half-rate, memory-4, maximum free distance NSC code having the generator polynomials of $g_0 = 1 + D + D^2 + D^4$, and $g_1 = 1 + D^3 + D^4$. The extrinsic information transfer functions of the inner code and the outer NSC code are shown in Fig. 10.19. It can be seen that the minimum convergence SNR threshold for the benchmarker system using the NSC outer code is about 1.2 dB, although we note that these curves are based on the assumption of having an infinite interleaver length and a Gaussian Log Likelihood Ratio (LLR) distribution. In the case of short block lengths, the actual SNR convergence threshold might be higher.

Hence, when constructing the IRCC, we choose the target inner code transfer function $T_{II}(i)$ at E_b/N_0 = 1.5 dB, and the number of iterations $F = 6$. For the constituent subcodes, we use those proposed in [345] except that code rates of $r_k > 0.75$ are excluded from our design for the sake of avoiding significant error floors. The resultant code rates of the subcodes span the range of $r_1 = 0.1, r_2 = 0.15, \cdots, r_{14} = 0.75$.

Initially the source constraint of eq. (10.8) was not imposed. By using the optimization procedure of [345], we arrive at the weight vector of $\boldsymbol{\alpha}_0 = [0\ 0\ 0\ 0\ 0.01\ 0.13\ 0.18\ 0.19\ 0.14\ 0.12\ 0.10\ 0.01\ 0.03\ 0.10]^T$, and the percentage of the input speech data bits protected by the different subcodes becomes $[0,\ 0,\ 0,\ 0,\ 0.6\%,\ 9.0\%,\ 14.4\%,\ 16.7\%,\ 14.0\%,\ 13.0\%,\ 11.5\%,\ 1.6\%,\ 4.2\%,\ 15.0\%]^T$. The extrinsic output of Decoder I after 6 iterations becomes $\mu_6 = 0.98$.

Observe in the context of the vector containing the corresponding speech bit fractions that only 0.6% of the source bits are protected by the $r_5 = 0.3$-rate subcode, whereas a total of 23.4% of the speech bits is protected by the $r_6 = 0.35$ and $r_7 = 0.4$-rate subcodes. In order to enhance the protection of the more sensitive speech bits, we impose now the source constraint of eq. (10.8) by requiring all the header information bits in a

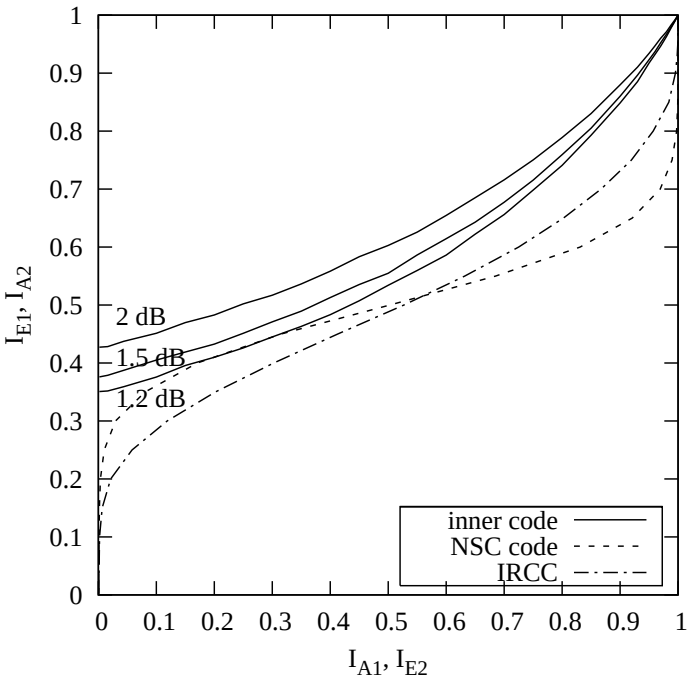

Figure 10.19: Extrinsic information transfer functions of the outer NSC code and the designed IRCC, as well as those of the inner code at $E_b/N_0 = 1.2$, 1.5 and 2 dB.

speech frame to be protected by the relatively strong $r_5 = 0.3$-rate subcode. More explicitly, we impose the constraint of $\alpha_5 r_5/0.5 \geq 7\%$, resulting in a new weight vector of $\boldsymbol{\alpha}_1 = [0\ 0\ 0\ 0\ 0.12\ 0.06\ 0.14\ 0.16\ 0.13\ 0.12\ 0.10\ 0.02\ 0.04\ 0.11]^T$, and the new vector of speech bit fractions becomes $[0,\ 0,\ 0,\ 0,\ 7.1\%,\ 4.0\%,\ 10.9\%,\ 14.8\%,\ 13.5\%,\ 13.3\%,\ 12.2\%,\ 2.7\%,\ 5.5\%,\ 16\%]^T$. The extrinsic output after 6 iterations is now slightly reduced to $\mu_6 = 0.97$, which is close to the maximum value of 0.98. Furthermore, now, 14.9% of the speech bits is protected by the $r_6 = 0.35$ and $r_7 = 0.4$-rate subcodes.

The extrinsic information transfer function of this IRCC is also shown in Fig. 10.19. As seen from the EXIT chart, the convergence SNR threshold for the system using the IRCC is lower than 1.2 dB and there is a wider EXIT chart tunnel between the inner code's curve and the outer code's curve which is particularly so at the low I_A values routinely encountered during the first couple of iterations. Hence, given a limited number of iterations, we would predict that the system using the IRCC may be expected to perform better than that using the NSC outer code in the range of E_b/N_0 = 1.5∼2 dB.

10.9.6 UEP AMR IRCC Performance Results

Finally, the achievable system performance was evaluated for a $K = 340$ speech bit per 20 ms transmission frame, resulting in an interleaver length of $L = 688$ bits, including 8 tail bits. This wideband-AMR speech coded [347] frame was generated at a bit rate of

15.85 kbps in the codec's mode 4. Before channel encoding, each frame of speech bits is rearranged according to the descending order of the error sensitivity of the bits by considering Figure 10.17, so that the more important data bits are protected by stronger IRCC subcodes. An S-random interleaver [352] was employed with $S = 15$, where all of the subcodes' bits are interleaved together, and 10 iterations were performed by the iterative decoder.

The BER performance of the UEP system using IRCCs and that of the Equal Error Protection (EEP) benchmarker system using the NSC code are depicted in Fig. 10.20. It can be seen that the UEP system outperforms the EEP system in the range of $E_b/N_0 = 1.5 \sim 2.5$ dB, which matches our performance prediction inferred from the EXIT chart analysis of Section 10.9.4.

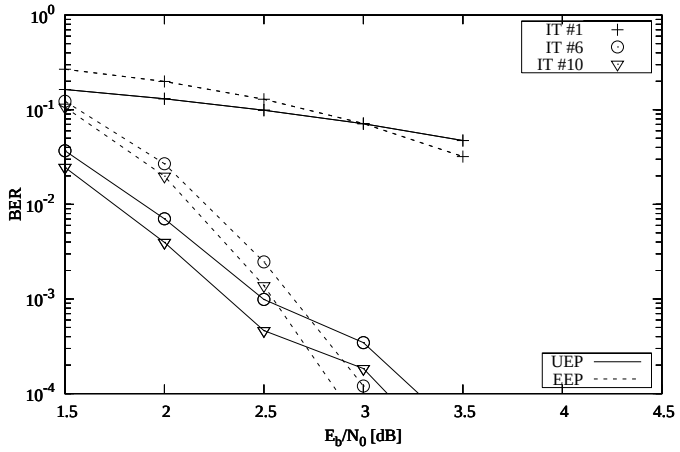

Figure 10.20: BER performance of both the UEP system employing the IRCC and the EEP system using the NSC code

The actual decoding trajectories of both the UEP system and the EEP system recorded at E_b/N_0=1.5 and 2 dB are shown in Fig. 10.21 and Fig. 10.22, respectively. These are obtained by measuring the evolution of mutual information at the input and output of both the inner decoder and the outer decoder as the iterative decoding algorithm is simulated. Due to the relatively short interleaver block length of 688 bits, the actual decoding trajectories do not closely follow the transfer functions especially when increasing the number of iterations. Nonetheless, the UEP system does benefit from having a wider open tunnel during the first couple of iterations and hence it is capable of reaching a higher extrinsic output in the end, resulting in a lower BER.

The BER profiles of the UEP system at E_b/N_0=1.5, 2 and 2.5 dB are plotted in Fig. 10.23. As intended, different fractions of the speech frame benefitted from different degrees of IRCC-aided protection. The first 60 bits represent the header information bits and the most sensitive speech bits, which require the lowest BER.

The SegSNR performances of both the UEP and EEP system are depicted in Fig. 10.24. The UEP system is seen to outperform the EEP system at $E_b/N_0 \leq 2.5$ dB. Above this E_b/N_0 point, the two systems attained almost the same SegSNRs. To achieve a good speech

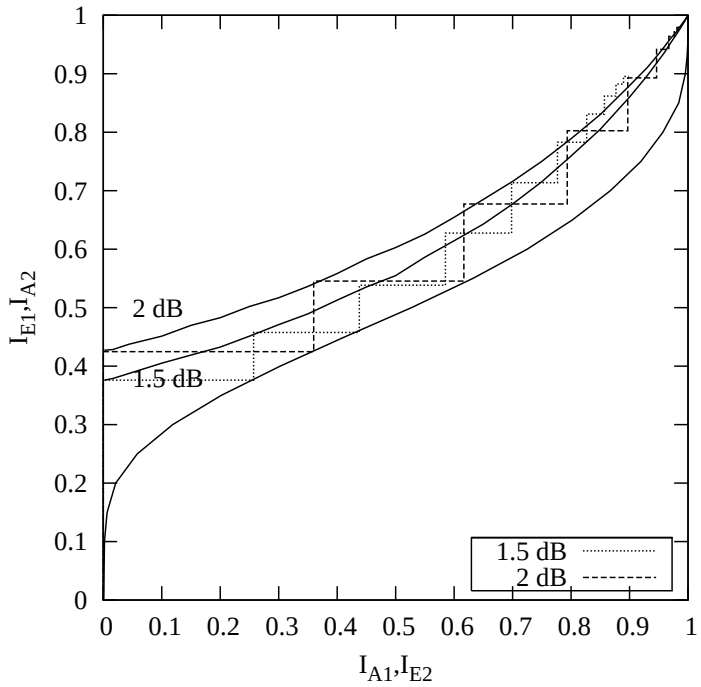

Figure 10.21: The EXIT chart and the simulated decoding trajectories of the UEP system using our IRCC as the outer code and a rate-1 recursive code as the inner code at both E_b/N_0=1.5 and 2 dB

quality associated with SegSNR>9 dB, the UEP system requires $E_b/N_0 \geq 2$ dB, about 0.3 dB less than the EEP system.

10.9.7 UEP AMR Conclusions

In Figure 10.17 of Sectionsec:amr-wb-codec we briefly exemplified the error sensitivity of the AMR-WB codec and then investigated the application of IRCCs for the sake of providing UEP for the AMR-WB speech codec. The IRCCs were optimized with the aid of EXIT charts and the design procedure used was illustrated with the aid of an example.

In the design of IRCCs, we aimed for matching the extrinsic information transfer function of the outer IRCC to that of the inner code, where that of the latter is largely determined by the channel SNR. At the same time, we imposed certain source constraints determined by the error sensitivity of the AMR-WB source bits. Hence the design method proposed here may be viewed as an attractive joint source/channel codec optimization.

The concatenated system using an IRCC benefits from having a low convergence SNR threshold. Owing to its design flexibility, various transfer functions can be obtained for an IRCC. We have shown that our IRCC was capable of achieving better convergence than a regular NSC code having the same constraint length and code rate. Hence the system using IRCCs has the potential of outperforming the corresponding arrangement using regular NSC

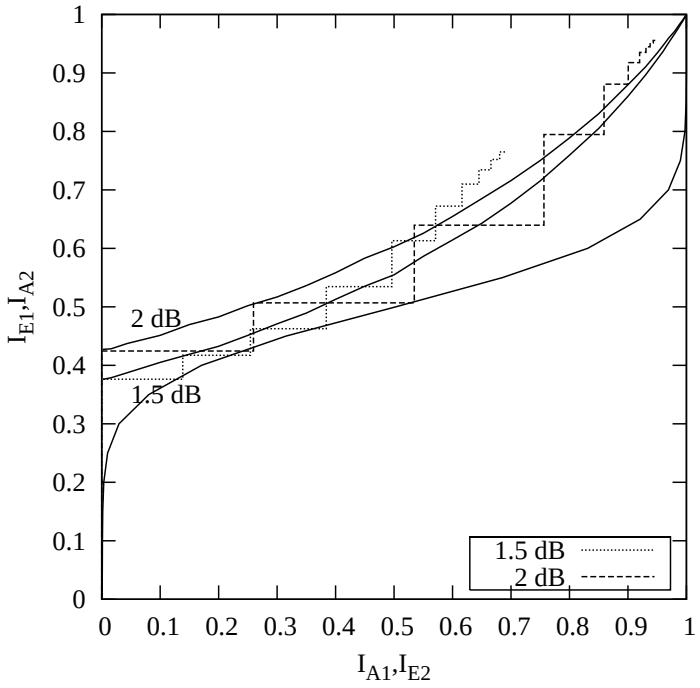

Figure 10.22: The EXIT chart and the simulated decoding trajectories of the EEP system using our NSC code as the outer code and a rate-1 recursive code as the inner code at both E_b/N_0=1.5 and 2 dB

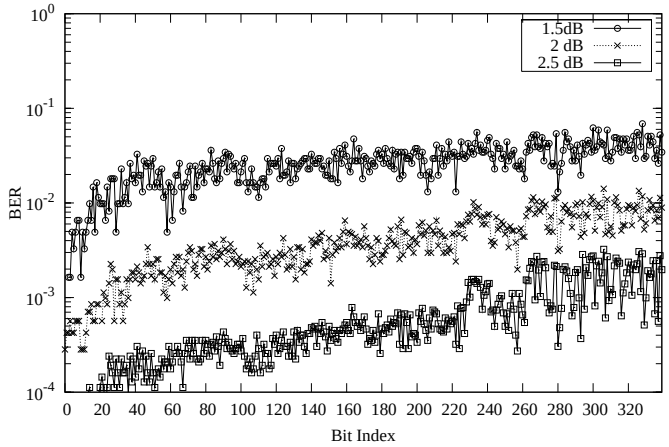

Figure 10.23: Bit error rate of the different speech bits after ten iterations at both E_b/N_0=1.5, 2 and 2.5dB recorded by transmitting 10^5 speech frames.

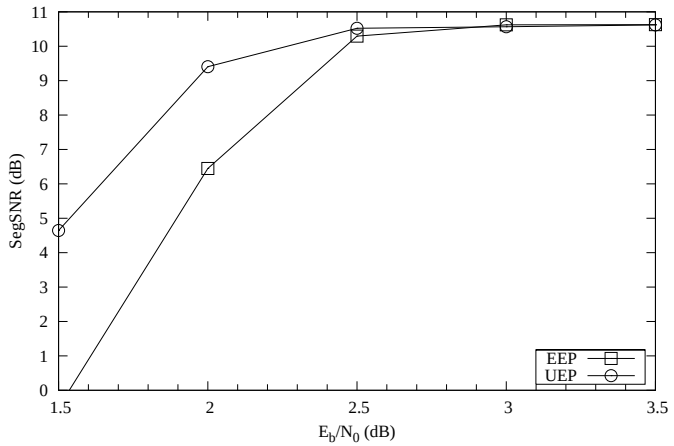

Figure 10.24: Comparison of SegSNRs of the AMR-WB speech codec using both EEP and UEP

codes in the low SNR region.

Furthermore, IRCCs are capable of providing UEP, since it is constituted by various sub-codes having different code rates and hence different error protection capabilities. Multimedia source information, such as speech, audio and video source can benefit from this property, when carefully designing the IRCC to match the source's bit sensitivity. Our future research aims for exchanging soft speech bits between the speech and channel decoders.

It is worth noting that an ISI channel can also be viewed as a rate-1 convolutional code, and the transfer function of an equalizer for a precoded ISI channel [353] is similar to that of the inner code here. Hence the proposed design method can be easily extended to ISI channels.

10.10 Chapter Summary

In Section 10.2 the various components of the narrowband AMR codec have been discussed. The error sensitivity of the narrowband AMR speech codec was characterised in 10.3, in order to match various channel codecs to the different-sensitivity bits of the speech codec. Specifically, we have shown that some bits in both the narrow- and wideband AMR codec are more sensitive to channel errors than others and hence require different grade of protection by channel coding. The error propagation properties of different bits over consecutive speech frames have also been characterized. We have shown how the degradations produced by errors propagate from one speech frame to the other and hence may persist over consecutive speech frames, especially in the scenario when the LSFs or the adaptive codebook delay bits were corrupted.

In Section 10.4, a joint-detection assisted near-instantaneously adaptive CDMA speech transceiver was designed, which allows us to switch between a set of different source and channel codec modes as well as transmission parameters, depending on the overall instantaneous channel quality. The 4.75 kbit/s and 10.2 kbit/s speech modes of the AMR codec have been employed in conjunction with the novel family of RRNS based channel coding, using

the reconfigurable BPSK or 4QAM based JD-CDMA scheme. In Section 10.6.2, the speech bits were mapped into three different protection classes according to their respective error sensitivities. In Section 10.8 the benefits of the multimode speech transceiver clearly manifested themselves in terms of supporting unimpaired speech quality under time-variant channel conditions, where a fixed-mode transceiver's quality would become severely degraded by the channel effects. The benefits of our dual-mode transceiver were further demonstrated with the aid of the characteristic waveforms displayed in Figure 10.15 and 10.16. Our AQAM/JD-CDMA scheme achieved the best compromise between unimpaired error-free speech quality and robustness, which has been verified by our informal listening tests shown in Table 10.6.

In Sectionsec:amr-wb-codec the wideband AMR codec was investigated and in Figure 10.17 we briefly exemplified the error sensitivity of the AMR-WB codec. Then IRCCs were invoked for the sake of providing UEP for the AMR-WB speech codec, which were optimized with the aid of the novel tools of EXIT charts. More specifically, we aimed for matching the EXIT transfer function of the outer IRCC to that of the inner code and we additionally imposed certain source constraints determined by the error sensitivity of the AMR-WB source bits. This design procedure may be readily extended to other joint source and channel coding schemes for the sake of attaining a near-capacity performance.

Chapter 11

MPEG-4 Audio Compression and Transmission

H-T. How and L. Hanzo

11.1 Overview of MPEG-4 Audio

The Moving Picture Experts Group (MPEG) was first established by the International Standard Organisation (ISO) in 1988 with the aim of developing a full audio-visual coding standard referred to as MPEG-1 [30–32]. The audio-related section MPEG-1 was designed to encode digital stereo sound at a total bit rate of 1.4 to 1.5 Mbps - depending on the sampling frequency, which was 44.1 kHz or 48 kHz - down to a few hundred kilobits per second [33]. The MPEG-1 standard is structured in layers, from Layer I to III. The higher layers achieve a higher compression ratio, albeit at an increased complexity. Layer I achieves perceptual transparency, i.e. subjective equivalence with the uncompressed original audio signal at 384 kbit/s, while Layer II and III achieve a similar subjective quality at 256 kbit/s and 192 kbit/s, respectively [34–38].

MPEG-1 was approved in November 1992 and its Layer I and II versions were immediately employed in practical systems. However, the MPEG Audio Layer III, MP3 for short only became a practical reality a few years later, when multimedia PCs were introduced having improved processing capabilities and the emerging Internet sparked off a proliferation of MP3 compressed teletraffic. This changed the face of the music world and its distribution of music. The MPEG-2 backward compatible audio standard was approved in 1994 [39], providing an improved technology that would allow those who had already launched MPEG-1 stereo audio services to upgrade their system to multichannel mode, optionally also supporting a higher number of channels at a higher compression ratio. Potential applications of the multichannel mode are in the field of quadraphonic music distribution or cinemas. Furthermore, lower sampling frequencies were also incorporated, which include 16, 22.05, 24, 32, 44.1 and 48 kHz [39]. Concurrently, MPEG commenced research into even higher-compression schemes,

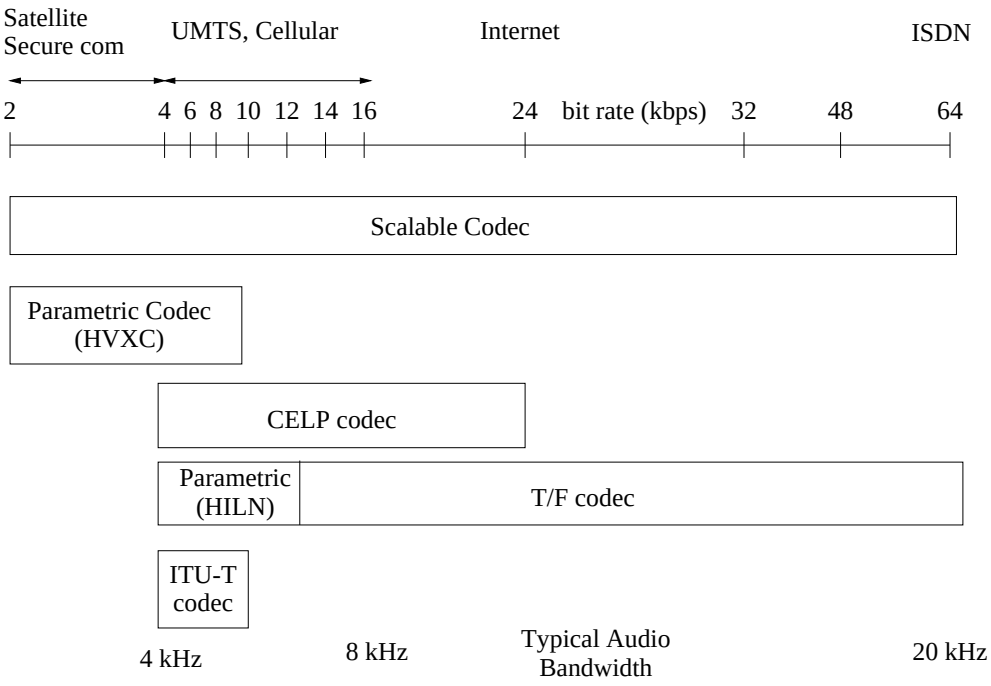

Figure 11.1: MPEG-4 framework [41].

relinquishing the backward compatibility requirement, which resulted in the MPEG-2 Advanced Audio Coding standard (AAC) standard in 1997 [40]. This provides those who are not constrained by legacy systems to benefit from an improved multichannel coding scheme. In conjunction with AAC, it is possible to achieve perceptual transparent stereo quality at 128 kbit/s and transparent multichannel quality at 320 kbit/s for example in cinema-type applications.

The MPEG-4 audio recommendation is the latest standard completed in 1999 [41–45], which offers in addition to compression further unique features that will allow users to interact with the information content at a significant higher level of sophistication than is possible today. In terms of compression, MPEG-4 supports the encoding of speech signals at bit rates from 2 kbit/s up to 24 kbit/s. For coding of general audio, ranging from very low bit rates up to high quality, a wide range of bit rates and bandwidths are supported, ranging from a bit rate of 8 kbit/s and a bandwidth below 4 kHz to broadcast quality audio, including monoaural representations up to multichannel configuration.

The MPEG-4 audio codec includes coding tools from several different encoding families, covering parametric speech coding, CELP-based speech coding and Time/Frequency (T/F) audio coding, which are characterised in Figure 11.1. It can be observed that a parametric coding scheme, namely Harmonic Vector eXcitation Coding (HVXC) was selected for covering the bit rate range from 2 to 4 kbit/s. For bit rates between 4 and 24 kbit/s, a CELP-coding scheme was chosen for encoding narrowband and wideband speech signals. For encoding

general audio signals at bit rates between 8 and 64 kbit/s, a time/frequency coding scheme based on the MPEG-2 AAC standard [40] endowed with additional tools is used. Here, a combination of different techniques was established, because it was found that maintaining the required performance for representing speech and music signals at all desired bit rates cannot be achieved by selecting a single coding architecture. A major objective of the MPEG-4 audio encoder is to reduce the bit rate, while maintaining a sufficiently high flexibility in terms of bit rate selection. The MPEG-4 codec also offers other new functionalities, which include bit rate scalability, object-based of a specific audio passage for example, played by a certain instrument representation, robustness against transmission errors and supporting special audio effects.

MPEG-4 consists of Versions 1 and 2. Version 1 [41] contains the main body of the standard, while Version 2 [46] provides further enhancement tools and functionalities, that includes the issues of increasing the robustness against transmission errors and error protection, low-delay audio coding, finely grained bit rate scalability using the Bit-Sliced Arithmetic Coding (BSAC) tool, the employment of parametric audio coding, using the CELP-based silence compression tool and the 4 kbit/s extended variable bit rate mode of the HVXC tool. Due to the vast amount of information contained in the MPEG-4 standard, we will only consider some of its audio compression components, which include the coding of natural speech and audio signals. Readers who are specifically interested in text-to-speech synthesis or synthetic audio issues are referred to the MPEG-4 standard [41] and to the contributions by Scheirer *et al.* [47, 48] for further information. Most of the material in this chapter will be based on an amalgam of References [34–38, 40, 41, 43, 44, 46, 49]. In the next few sections, the operations of each component of the MPEG-4 audio component will be highlighted in greater detail. As an application example, we will employ the Transform-domain Weighted Interleaved Vector Quantization (TWINVQ) coding tool, which is one of the MPEG-4 audio codecs in the context of a wireless audio transceiver in conjunction with space-time coding [50] and various Quadrature Amplitude Modulation (QAM) schemes [51]. The audio transceiver is introduced in Section 11.5 and its performance is discussed in Section 11.5.6.

11.2 General Audio Coding

The MPEG-4 General Audio (GA) coding scheme employs the Time/Frequency (T/F) coding algorithm, which is capable of encoding music signals at bit rates from 8 kbit/s per channel and stereo audio signals at rates from 16 kbit/s per stereo channel up to broadcast quality audio at 64 kbit/s per channel and higher. This coding scheme is based on the MPEG-2 Advanced Audio Coding (AAC) standard [40], enriched by further addition of tools and functionalities. The MPEG-4 GA coding incorporates a range of state-of-the-art coding techniques, and in addition to supporting fixed bit rates it also accommodates a wide range of bit rates and variable rate coding arrangements. This was facilitated with the aid of the continuous development of the key audio technologies throughout the past decades. Figure 11.2 shows in an non-exhaustive fashion some of the important milestones in the history of perceptual audio coding, with emphasis on the MPEG standardization activities. These important developments and contributions, which will be highlighted in more depth during our further discourse throughout this chapter, have also resulted in several well-known commercial audio coding standards, such as the Dolby AC-2/AC-3 [412], the Sony Adaptive Transform Acous-

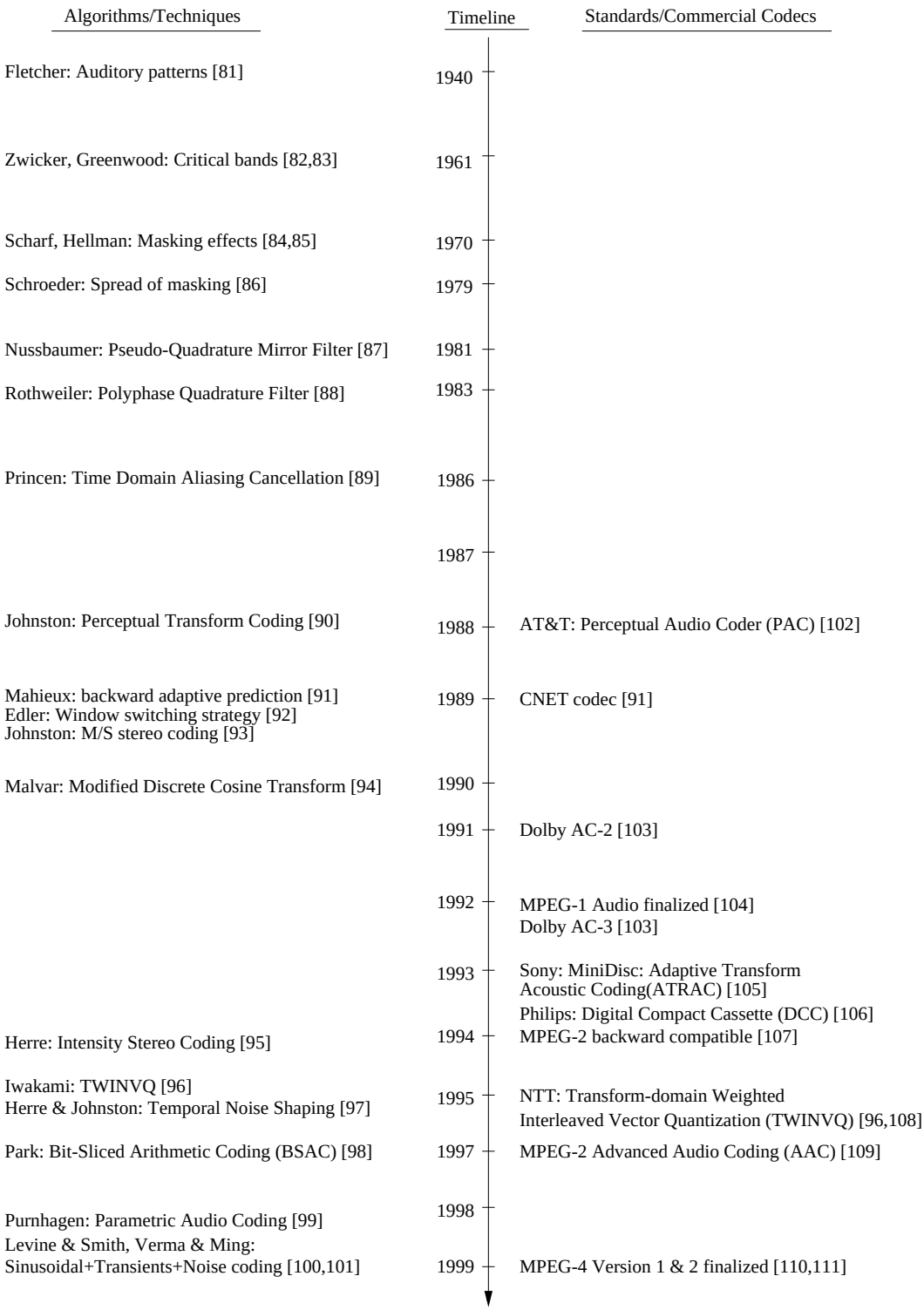

Figure 11.2: Important milestones in the development of perceptual audio coding.

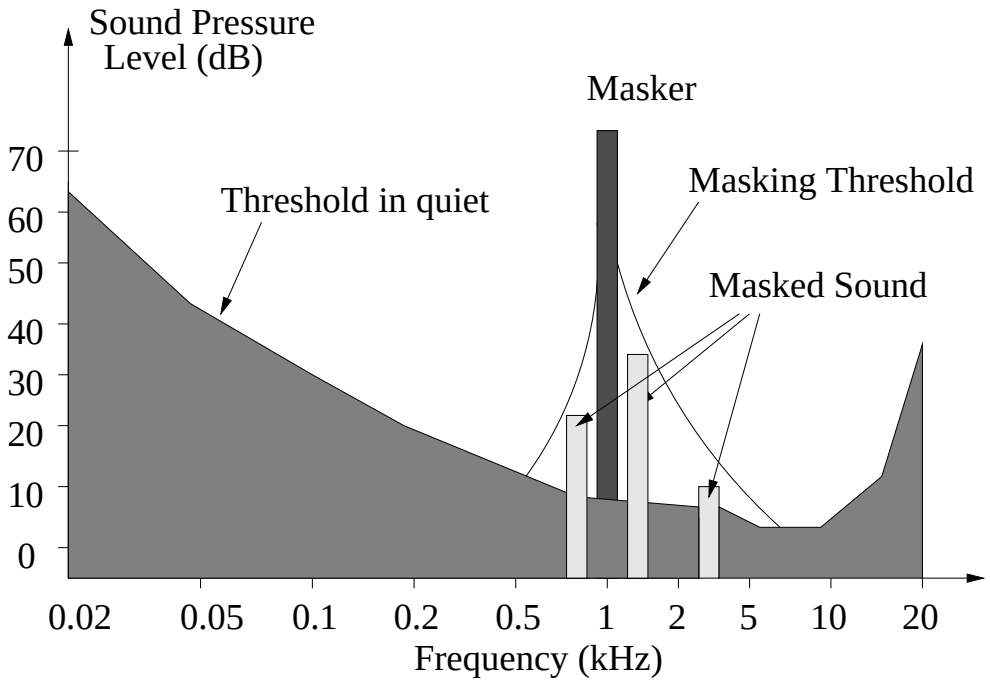

Figure 11.3: Threshold in quiet and masking threshold [416].

tic Coding (ATRAC) for MiniDisc [413], the Lucent Perceptual Audio Coder (PAC) [414] and Philips Digital Compact Cassette (DCC) [415] algorithms. Advances in audio bit rate compression techniques can be attributed to four key technologies:

A) Perceptual Coding

Audio coders reduce the required bit rates by exploiting the characteristics of masking the effects of quantization errors in both the frequency and time domains by the human auditory system, in order to render its effects perceptually inaudible [417–420]. The foundations of modern auditory masking theory were laid down by Fletcher's seminal paper in 1940 [421]. Fletcher [421] suggested that the auditory system behaves like a bank of bandpass filters having continuously overlapping passbands. Research has shown that the ear appears to perceive sounds in a number of critical frequency bands, as shown by Zwicker [418] and Greenwood [422]. This model of the ear can be roughly described as a bandpass filterbank, consisting of overlapping bandpass filters having bandwidths on the order of 100 Hz for signal frequencies below 500 Hz. By contrast, the bandpass filter bandwidths of this model may be as high as 5000 Hz at high frequencies. There exists up to twenty five such critical bands in the frequency range up to 20 kHz [418]. Auditory masking refers to the mechanism by which a fainter, but distinctly audible signal becomes inaudible, when a louder signal occurs simultaneously (simultaneous masking), or within a very short time (forward or backward masking) [423]. More specifically, in the case of simultaneous masking the two sounds occur at the same time, for example in a scenario, when a conversation (masked signal) is rendered

inaudible by a passing train (the masker). Forward masking is encountered when the masked signal remains inaudible for a time after the masker has ended, while an example of this phenomenon in backward masking takes place when the masked signal becomes inaudible even before the masker begins. An example is the scenario during abrupt audio signal attacks or transients, which create a pre-and post-masking regions in time during which a listener will not be able to perceive signals beneath the audibility thresholds produced by a masker. Hence, specific manifestation of masking depends on the spectral composition of both the masker and masked signal, and their variations as a function of time [424]. Important conclusions, which can be drawn from all three masking scenarios [424, 425] are firstly, that simultaneous masking is more effective when the frequency of the masked signal is equal to or higher than that of the masker. This result is demonstrated in Figure 11.3, where a masker rendered three masked signals inaudible, which occurred at both lower and higher frequencies than the masker. Secondly, while forward masking is effective for a considerable time after the masker has decayed, backward masking may only be effective for less than 2 or 3 ms before the onset of the masker [424].

A *masking threshold* can be determined, whereby signals below this threshold will be inaudible. Again, Figure 11.3 depicts an example of the masking threshold of a narrowband masker, having three masked signals in the neighbourhood. As long as the sound pressure levels of the three maskees are below the masking threshold, the corresponding signals will be masked. Observe that the slope of the masking threshold is steeper towards lower frequencies, which implies that higher frequencies are easier to mask. When no masker is present, a signal will be inaudible if its sound pressure level is below the *threshold in quiet*, as displayed in Figure 11.3. The *threshold in quiet* characterizes the amount of energy required for a pure tone to be detectable by a listener in a noiseless environment. The situation discussed here only involved one masker, but in real life, the source signals may consists of many simultaneous maskers, each having its own masking threshold. Thus, a *global masking threshold* has to be computed, which describes the threshold of *just noticeable distortions* as a function of frequency [424].

B) Frequency Domain Coding

The evolution of time/frequency mapping or filterbank based techniques has contributed to the rapid development in the area of perceptual audio coding. Some of the earliest frequency domain audio coders include contributions from Brandenburg [426] and Johnston [427] although subband based narrow- and wideband speech codecs were developed during the late 1970s and early 1980s [428–430]. Frequency domain encoders [431, 432], which are employed in all MPEG codecs offer a convenient way of controlling the frequency-domain distribution of the quantization noise, in conjunction with dynamic bit allocation applied to the quantization of subband signals or transform coefficients. Essentially, the filterbank divides the spectrum of the input signal into frequency subbands, which host the contributions of the fullband signal in the subband concerned. Given the knowledge of an explicit perceptual model, the filterbank facilitates the task of perceptually motivated noise shaping and that of identifying the perceptually unimportant subbands. It is important to choose the appropriate filterbank for bandsplitting. An adaptive filterbank exhibiting time-varying resolutions in both the time and frequency domain is highly desirable. This issue has motivated intensive research, experimenting with various switched or hybrid filterbank structures, where the switching decisions were based on the time-variant input signal characteristics [433].

Depending on the frequency domain resolution, we can categorize frequency domain coders

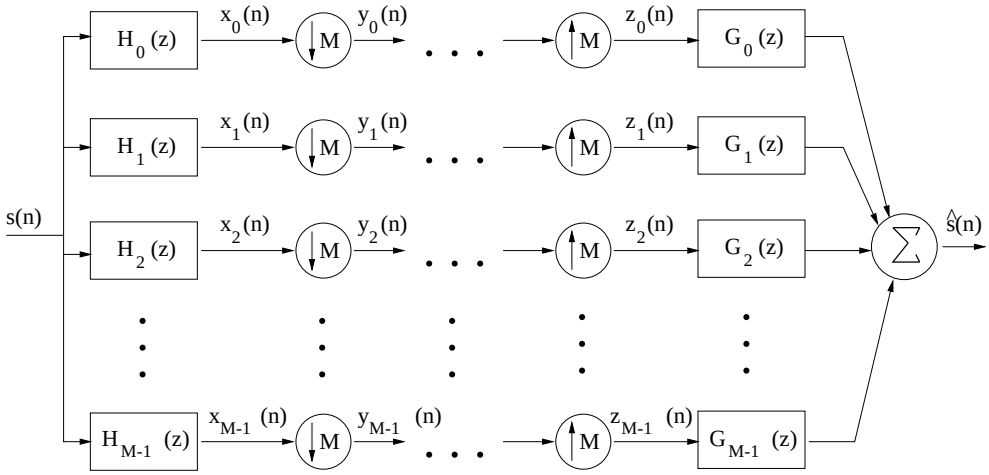

Figure 11.4: Uniform M-band analysis-synthesis filterbank [420].

as either transform coders [426, 427], or subband coders [434–436]. The basic principle of transform coders is the multiplication of overlapping blocks of audio samples with a smooth time-domain window function, followed by either the Discrete Fourier Transform (DFT) or the Discrete Cosine Transform (DCT) [437], which transform the input time-domain signal into a high resolution frequency domain representation, consisting of nearly uncorrelated spectral lines or transform coefficients. The transform coefficients are subsequently quantized and transmitted over the channel. At the decoder, the inverse transformation is applied. By contrast, in subband codecs, the input signal is split into several uniform or non-uniform width subbands using critically sampled [435], Perfect Reconstruction [438] (PR) or non-PR [439] filterbanks. For example, as shown in Figure 11.4, when an input signal is split into M bandpass signals, critical decimation by a factor of M is applied. This means that every mth sample of each bandpass signal is retained, which ensures that the total number of samples across the subbands equals the number of samples in the original input signal. At the synthesis stage, a summation of the M bandpass signals is performed, which leads to interpolation between samples at the output.

The traditional categorization into the families of subband and transform coders has been blurred by the emerging trend of combining both techniques in the codec design, as exemplified by the MPEG codecs, which employ both techniques. In the contribution by Temerinac [440], it was shown mathematically that all transforms used today in the audio coding systems can be viewed as filterbanks. All uniform-width subband filterbanks can be viewed as transforms of splitting a full-band signal into n components [440]. One of the first filterbank structure proposed in early 1980s, was based on Quadrature Mirror Filters (QMF) [428]. Specifically, a near-PR QMF filter was proposed by Nussbaumer [441] and Rothweiler [439]. In order to derive the pseudo-QMF structure, firstly the analysis-by-synthesis filters have to meet the mirror image condition of [439]:

$$g_k(n) = h_k(L-1-n) \quad . \tag{11.1}$$

Additionally, the precise relationships between the analysis and synthesis filters h_k and g_k have to be established in order to eliminate aliasing. With reference to Figure 11.4, the analysis and synthesis filters which eliminate both aliasing and phase distortions are given by [435] :

$$h_k(n) = 2w(n)cos\left[\frac{\pi}{M}(k+0.5)\left(n-\frac{(L-1)}{2}\right)+\theta_k\right] \tag{11.2}$$

and

$$g_k(n) = 2w(n)cos\left[\frac{\pi}{M}(k+0.5)\left(n-\frac{(L-1)}{2}\right)-\theta_k\right] \tag{11.3}$$

respectively, where

$$\theta_k = (-1)^k\frac{\pi}{4} \quad . \tag{11.4}$$

The filterbank design is now reduced to the design of the time-domain window function, $w(n)$. The principles of Pseudo-QMFs have been applied in both the MPEG-1 and MPEG-2 schemes, which employ a 32-channel Pseudo-QMF for implementing spectral decomposition in both the Layer I and II schemes. The same Pseudo-QMF filter was used in conjunction with a PR cosine-modulated filterbank in Layer III in order to form a hybrid filterbank [35]. This hybrid combination could provide a high frequency resolution by employing a cascade of a filterbank and an Modified Discrete Cosine Transform (MDCT) transform that splits each subband further in the frequency domain [37].

The MDCT [437], which has been defined in the current MPEG-2 and 4 codecs, was first proposed under the name of Time Domain Aliasing Cancellation (TDAC) by Princen and Bradley [442] in 1986. It is essentially a PR cosine modulated filterbank satisfying the constraint of $L = 2M$, where L is the window size while M is the transform length. In conventional block transforms, such as the DFT or DCT, blocks of samples are processed independently, due to the quantization errors the decoded signal will exhibit discontinuities at the block boundaries since in the context of conventional block-based transforms the time-domain signal is effectively multiplied by a rectangular time-domain window, its sinc-shaped frequency domain representation is convolved with the spectrum of the audio signal. This results in the well-known Gibbs phenomenon. This problem is mitigated by applying the MDCT, using a specific window function in combination with overlapping the consecutive time-domain blocks. As shown in Figure 11.5, a window of $2M$ samples collected from two consecutive time-domain blocks undergoes cosine transformation, which produces M frequency-domain transform coefficients. The time-domain window is then shifted by M samples for computing the next M transform coefficients. Hence, there will be a 50% overlap in each consecutive DCT transform coefficient computation. This overlap will ensure a more smooth evolution of the reconstructed time-domain samples, even though there will be some residual blocking artifacts due to the quantization of the transform coefficients. Nonetheless, the MDCT virtually eliminates the problem of blocking artifacts that plague the reconstructed signal produced by non-overlapped transform coders. This problem often manifestated itself as a periodic clicking in the reconstructed audio signals. Again, the processes associated with the MDCT-based overlapped analysis and the corresponding overlap-add synthesis are

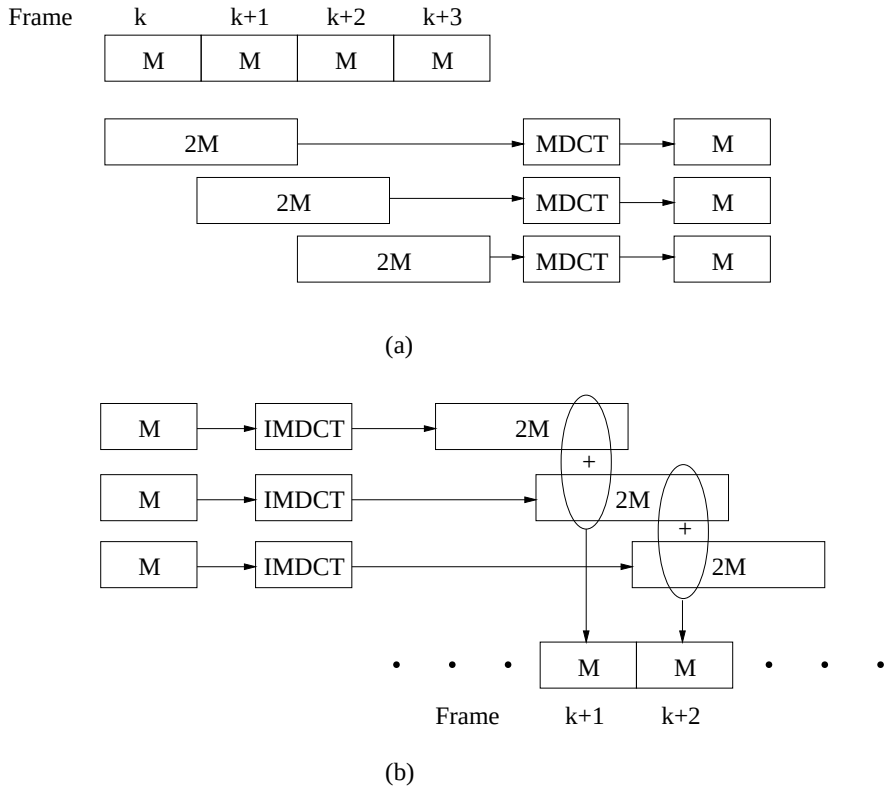

Figure 11.5: (a) MDCT analysis process, $2M$ samples are mapped into M spectral coefficients (b) MDCT synthesis process, M spectral coefficients are mapped to a vector of $2M$ samples which is overlapped by M samples with the vector of $2M$ samples from the previous frame, and then added together to obtain the reconstructed output of M samples [420].

illustrated in Figure 11.5. At the analysis stage, the forward MDCT is defined as [443]:

$$X(k) = \sum_{n=0}^{2M-1} x(n)h_k(n), \quad k = 0...M-1\ , \tag{11.5}$$

where the M MDCT coefficients $X(k), \quad k = 0...M-1$ are generated by computing a series of inner products between the $2M$ samples $x(n)$ of the input signal and the corresponding analysis filter impulse response $h_k(n)$. The analysis filter impulse response, $h_k(n)$, is given by [443]:

$$h_k(n) = w(n)\sqrt{\frac{2}{M}}cos\left[\frac{(2n+M+1)(2k+1)\pi}{4M}\right]\ , \tag{11.6}$$

where $w(n)$ is a window function, and the specific window function used in the MPEG stan-

dard is the sine window function, given by [443]:

$$w(n) = sin\left[\left(n + \frac{1}{2}\right)\frac{\pi}{2M}\right] \quad . \tag{11.7}$$

At the synthesis stage, the inverse MDCT is defined by [443]:

$$x(n) = \sum_{k=0}^{M-1} [X(k)h_k(n) + X^P(k)h_k(n+M)] \quad . \tag{11.8}$$

In Equation 11.8 we observe that the time-domain reconstructed sample $x(n)$ is obtained by computing a sum of the basis vectors $h_k(n)$ and $h_k(n+M)$ weighted by the transform coefficients $X(k)$ and $X^P(k)$ on the basis of the current and previous blocks as it was also illustrated in Figure 11.5. More specifically, the first M-sample block of the kth basis vector, $h_k(n)$, for $0 \leq n \leq M-1$, is weighted by the kth MDCT coefficients of the current block. By contrast, the second M-sample block of the kth basis vector, $h_k(n)$, for $M \leq n \leq 2M-1$ is weighted by the kth MDCT coefficients of the previous block, namely by $X^P(k)$. The inverse MDCT operation is also illustrated in Figure 11.5.

C) Window Switching

The window switching strategy was first proposed in 1989 by Edler [444], where a bit rate reduction method was proposed for audio signals based on overlapping transforms. More specifically, Edler proposed adapting the window functions and the transform lengths to the nature of the input signal. This improved the performance of the transform codec in the presence of impulses and rapid energy on-set occurrences in the input signal. The notion of applying different windows according to the input signal's properties has been subsequently incorporated in the MPEG codecs employing the MDCT, for example MPEG-1 Layer III and MPEG-2 AAC codecs [40].

Typically, a long time-domain window is employed for encoding the identifiable stationary signal segments while primarily a short window is used for localizing the pre-echo effects due to the occurrence of sudden signal on-sets, as experienced during transient signal periods, for example [40]. In order to ensure that the conditions of PR-based analysis and synthesis filtering are property are preserved, transitional windows are needed for switching between the long and short windows [443]. These transitional windows are depicted graphically in Figure 11.6, utilizing four window functions, namely long, short, start and stop windows, which are also used in the MPEG-4 General Audio coding standard.

D) Dynamic Bit Allocation

Dynamic bit allocation aims for assigning bits to each of the quantizers of the transform coefficients or subband samples, in such a way that the overall perceptual quality is maximized [445]. This is an iterative process, where in each iteration, the number of quantizing levels is increased, while satisfying the constraint that the number of bits used must not exceed the number of bits available for that frame.

Furthermore, another novel bit allocation technique referred to as the "bit reservoir" scheme was proposed for accommodating the sharp signal on-sets, which resulted in an increased number of required bits during the encoding of transient signals [445]. This is due to the fact that utilising the window switching strategy does not succeed in avoiding all audible pre-echos, in particular, when sudden signal on-set occurrences near the end of a transform

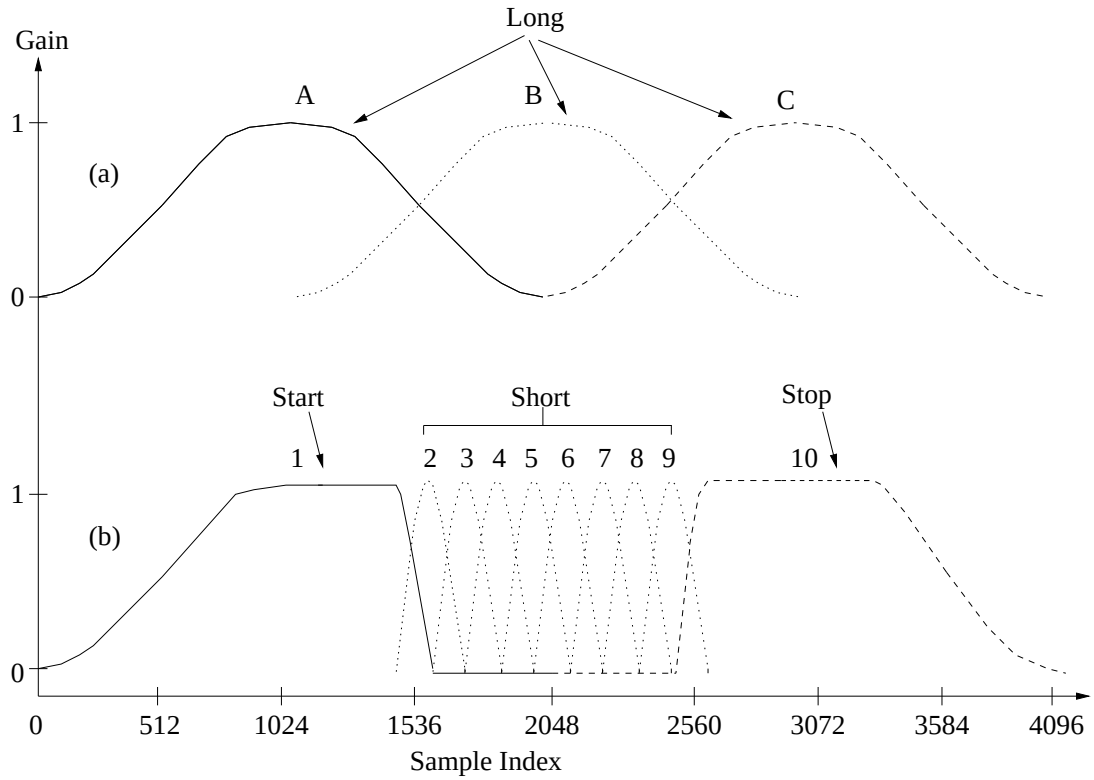

Figure 11.6: Window transition during (a) steady state using long windows and (b) transient conditions employing start, short, and stop windows [40].

block [420]. In block-based schemes like conventional transform codecs, the inverse transform spreads the quantization errors evenly in time over the duration of the reconstruction block. This results in audible unmasked distortion throughout the low-energy signal segment the instant of the signal attack [420]. Hence, the "bit reservoir" technique was introduced for allocating more bits to those frames, which invoked pre-echo control. This "bit reservoir" technique was employed in the MPEG Layer III and MPEG-2 AAC codecs [40].

11.2.1 Advanced Audio Coding

The MPEG-2 Advanced Audio Coding (AAC) scheme was declared an international standard by MPEG at the end of April 1997 [40]. The main driving factor behind the MPEG-2 AAC initiative was the quest for an efficient coding method for multichannel surround sound signals such as the 5-channel (left, right, centre, left-surround and right-surround) system designed for cinemas. The main block diagram of the MPEG-4 Time/Frequency (T/F) codec is as shown in Figure 11.7, which was defined to be backward compatible to the MPEG-2 AAC scheme [40].

In this section we commence with an overview of the AAC profiles based on Figure 11.7

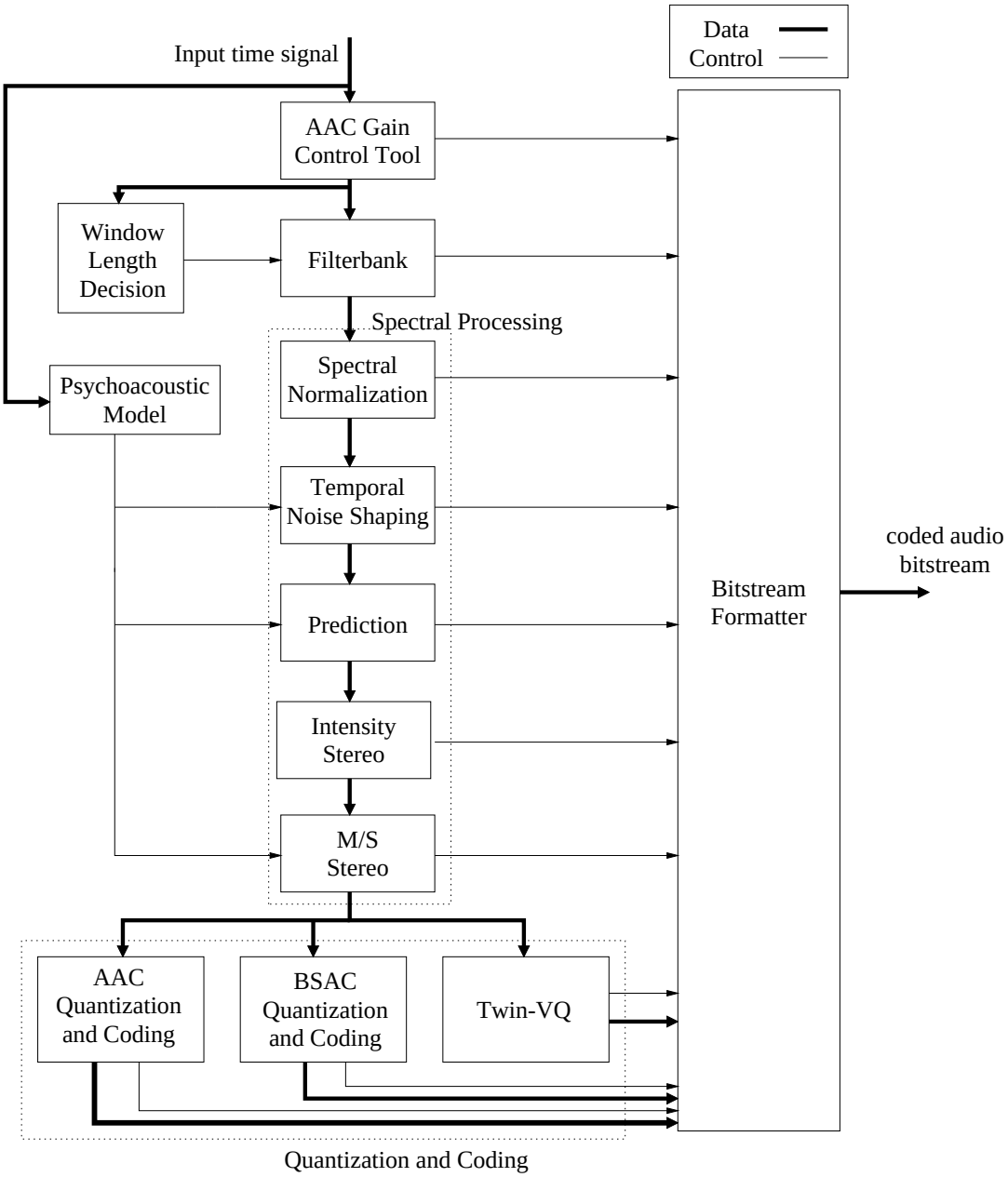

Figure 11.7: Block diagram of MPEG-4 T/F-based encoder [41].

and each block will be discussed in more depth in Section 11.2.2 - 11.2.10. Following the diagram shown in Figure 11.7, the T/F coder first decomposes the input signal into a T/F representation by means of an analysis filterbank prior to subsequent quantization and coding. The filterbank is based on the Modified Discrete Cosine Transform (MDCT) [442], which is also known as Modulated Lapped Transform (MLT) [446]. In the case when the Scalable Sampling Rate (SSR) mode is invoked, the MDCT will be preceded by a Polyphase Quadrature Filter (PQF) [439] and a gain control module, which are not explicitly shown in Figure 11.7 but will be described in Section 11.2.2. In the encoding process, the filterbank takes in a block of samples, applies the appropriate windowing function and performs the MDCT within the filterbank block. The MDCT block length can be either 2048 or 256 samples, switched dynamically depending on the input signal's characteristics. This window switching mechanism was first introduced by Edler in [444]. Long block transform processing (2048 samples) will improve the coding efficiency of stationary signals, but problems might be incurred when coding transients signals. Specifically, this gives rise to the problem of pre-echos, which occur when a signal exhibiting a sudden sharp signal envelope rise begins near the end of a transform block [420]. In block-based schemes, such as transform codecs, the inverse transform will spread the quantization error evenly in time over the reconstructed block. This may result in audible unmasked quantization distortion throughout the low-energy section preceding the instant of the signal attack [420]. By contrast, a shorter block length processing (256 samples) will be optimum for coding transient signals, although it suffers from inefficient coding of steady-state signals due to the associated poorer frequency resolution.

Figure 11.6 shows the philosophy of the block switching mechanisms during both steady state and transient conditions. Specifically, two different window functions, the Kaiser-Bessel derived (KBD) window [412] and the sine window can be used for windowing the incoming input signal for the sake of attaining an improved frequency selectivity and for mitigating the Gibb-oscillation, before the signal is transformed by the MDCT [412]. The potential problem of appropriate block alignment due to window switching is solved as follows. Two extra window shapes, so-called start and stop windows are introduced together with the long and short windows depicted in Figure 11.6. The long window consists of 2048 samples while a short window is composed of eight short blocks arranged to overlap by 50% with each other. At the boundaries between long and short blocks, half of the transform blocks overlap with the start and stop windows. Specifically, the *start* window enables the transition between the long and short window types. The left half of a *start* window seen at the bottom of Figure 11.6 shares the form as the left half of the long window type depicted at the top of Figure 11.6. The right half of the *start* window has the value of unity for one-third of the length and the shape of the right half of a short window for the central one-third duration of its total length, with remaining one-third of the *start* window duration length set to zero. Figure 11.6 (a) shows at the top of Figure 11.6 the steady state condition, where only long transform blocks are employed. By contrast, Figure 11.6 (b) displays the block switching mechanism, where we can observe that the start (#1) and stop (#10) window sequences ensure a smooth transition between long and short transforms. The start window can be either the KBD or the sine-window, in order to match the previous long window type, while the stop window is the time-reversed version of the start window.

Like all other perceptually motivated coding schemes, the MPEG-4 AAC-based codec makes use of the signal masking properties of the human ear, in order to reduce the required bit rate. By doing so, the quantization noise is distributed to frequency bands in such a

way that it is masked by the total signal and hence it remains inaudible. The input audio signal simultaneously passes through a psychoacoustic model as shown in Figure 11.7, that determines the ratio of the signal energy to the masking threshold. An estimate of the masking threshold is computed using the rules of psychoacoustics [34]. Here, a perceptual model similar to the MPEG-1 psychoacoustic model II [40] is used, which will be described in Section 11.2.3. A signal-to-mask ratio is computed from the masking threshold, which is used to decide on the bit allocation, in an effort to minimize the audibility of the quantization noise.

After the MDCT carried out in the filterbank block of Figure 11.7 the spectral coefficients are passed to the Spectral Normalization 'toolbox', if the TWINVQ mode is used. The Spectral Normalization tool will be described in Section 11.2.9. For AAC-based coding, the spectral coefficients will be processed further by the Temporal Noise Shaping (TNS) 'toolbox' of Figure 11.7, where TNS uses a prediction approach in the frequency domain for shaping and distributing the quantization noise over time.

Time domain 'Prediction' block of Figure 11.7 or Long-Term Prediction (LTP) is an important tool, which increases redundancy reduction of stationary signals. It utilises a second order backward adaptive predictor, which is similar to the scheme proposed by Mahieux [447]. In the case of multichannel input signals, 'Intensity Stereo' coding is also applied as seen in Figure 11.7, which is a method of replacing the left and right stereo signals by a single signal having embedded directional information. Mid/Side (M/S) stereo coding, as described by Johnston [448] can also be used as seen in Figure 11.7, where instead of transmitting the left and right signals, the sum and difference signals are transmitted.

The data-compression based bit rate reduction occurs in the quantization and coding stage, where the spectral values can be coded either using the AAC, Bit Sliced Arithmetic Coding [449] (BSAC) or TWINVQ [450] techniques as seen in Figure 11.7. The AAC quantization scheme will be highlighted in Section 11.2.6 while the BSAC and TWINVQ-based techniques will be detailed in Section 11.2.8 and 11.2.9, respectively. The AAC technique invokes an adaptive non-linear quantizer and a further noise shaping mechanism employing scale-factors is implemented. The allocation of bits to the spectral values is carried out according to the psychoacoustic model, with the aim of suppressing the quantization noise below the masking threshold. Finally, the quantized and coded spectral coefficients and control parameters are packed into a bitstream format ready for transmission. In the following sections, the individual components of Figure 11.7 will be discussed in further details.

11.2.2 Gain Control Tool

When the Scalable Sampling Rate (SSR) mode is activated, which facilitates the employment of different sampling rates, the MDCT transformation taking place in the Filterbank block of Figure 11.7 is preceded by uniformly-spaced 4-band Polyphase Quadrature Filter [441] (PQF), plus a gain control module [41]. The PQF splits the input signal into four frequency bands of equal width. When the SSR mode is invoked, lower bandwidth output signals, and hence lower sampling rate signals can be obtained by neglecting the signals residing in the lower-energy upper bands of the PQF. In the scenario, when the bandwidth of the input signal is 24 kHz, equivalent to a 48 kHz sampling rate, output bandwidths of 18, 12 and 6 kHz can be obtained when one, two or three PQF outputs are ignored, respectively [40].

The purpose of the gain control module is to appropriately attenuate or amplify the output

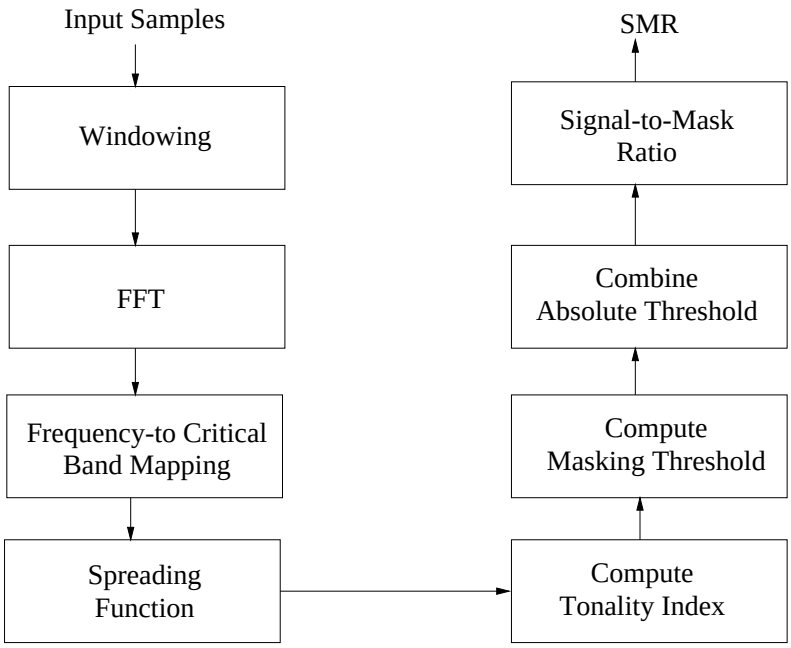

Figure 11.8: Flow diagram of the psychoacoustic model II in MPEG-4 AAC coding.

of each PQF band, in order to reduce the potential pre-echo effects [420]. The gain control module, which estimates and adjusts the gain factor of the subbands, according to the psychoacoustic requirements, can be applied independently to each subband. At the encoder, the gain control 'toolbox' receives the time domain signals as its input and outputs the gain control data and the appropriately scaled signal whose length is equal to the length of the MDCT window. The 'gain control data' consists of the number of bands which experienced gain modification, the number of modified segments and the indices indicating the location and level of gain modification for each segment. Meanwhile, the 'gain modifier' associated with each PQF band controls the gain of each band. This effectively smoothes the transient peaks in the time domain prior to MDCT spectral analysis. Subsequently, the normal procedure of coding stationary signals using long blocks can be applied.

11.2.3 Psychoacoustic Model

As argued in Section 11.2, the MPEG-4 audio codec and other perceptually optimized codecs reduce the required bit rate by taking advantage of the human auditory system's inability to perceive the quantization noise satisfying the conditions of auditory masking. Again, perceptual masking occurs, when the presence of a strong signal renders the weaker signals surrounding it in the frequency-domain imperceptible [424]. The psychoacoustic model used in the MPEG-4 audio codec is similar to the MPEG-1 psychoacoustic model II [34].

Figure 11.8 shows the flow chart of the psychoacoustic model II. First a Hann window [41]

is applied to the input signal and then the Fast Fourier Transform (FFT) provides the necessary time-frequency mapping. The Hann window is defined as [41]:

$$w(n) = \frac{1}{2}[1 - cos(\frac{2\pi n}{N})] \tag{11.9}$$

where N is the FFT length. This windowing procedure is applied for the sake of reducing the frequency-domain Gibbs oscillation potentially imposed by a rectangular transform window. Depending on whether the signal's characteristics are of stationary or transient nature, FFT sizes of either 1024 or 128 samples can be applied. The FFT-based spectral coefficient values are then grouped according to the corresponding critical frequency band widths. This is achieved by transforming the spectral coefficient values into the "partition index" domain, where the partition indices are related near-linearly to the critical bands that were summarised in Figure 11.9 (a) recorded at the sampling rate of 44.1 kHz. At low frequencies, a single spectral line constitutes a partition, while at high frequencies many lines will be combined in order to form a partition, as displayed in Figure 11.9 (b). This facilitates the appropriate representation of the critical bands of the human auditory system [36]. Tables of the mapping functions between the spectral and partition domains and their respective values for threshold in quiet are supplied in the MPEG-4 standard for all available sampling rates [41].

During the FFT process of Figure 11.8, the polar representation of the transform-domain coefficients is also calculated. Both the magnitude and phase of this polar representation will be used for the calculation of the 'predictability measure', which is used for quantifying the predictability of the signal, as an indicator of the grade of tonality. The psychoacoustic model identifies the tonal and noise-like components of the audio signal, because the masking abilities of the two types of signals differ. In this psychoacoustic model, the masking ability of a tone masking the noise, which is denoted by $TMN(b)$, is fixed at 18 dB in all the partitions, which implies that any noise within the critical band more than 18 dB below $TMN(b)$ will be masked by the tonal component. The masking ability of noise masking tone, which is denoted by $NMT(b)$, is set to 6 dB for all partitions. The previous two frequency-domain blocks are used for predicting the magnitude and phase of each spectral line for the current frequency-domain block, via linear interpolation in order to obtain the 'predictability' values for the current block. Tonal components are more predictable and hence will have higher tonality indices. Furthermore, a spreading function [41] is applied in order to take into consideration the masking ability of a given spectral component, which could spread across its surrounding critical band.

The masking threshold is calculated in Figure 11.8 by using the tonality index and the threshold in quiet, T_q, which is known as the lower threshold bound above which a sound is audible. The masking threshold in each frequency-domain partition corresponds to the power spectrum multiplied by an attenuation factor given by [41]:

$$Attenuation\_Factor = 10^{-SNR(b)}, \tag{11.10}$$

implying that the higher the SNR, the lower the attenuation factor and also the masking threshold, where the Signal-to-Noise (SNR) ratio is derived as:

$$SNR(b) = tb(b) \cdot TMN(b) + (1 - tb(b)) \cdot NMT(b), \tag{11.11}$$

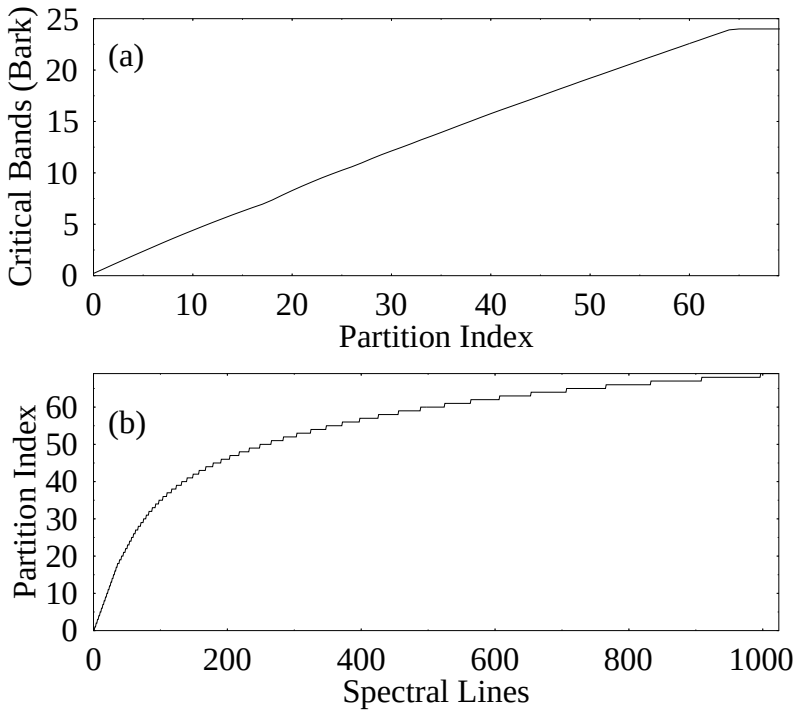

Figure 11.9: (a) Relationship between the partition index and critical bands. (b) The conversion from the FFT spectral lines to the partition index domain at the sampling rate of 44.1 kHz for a total of 1024 spectral lines per time-domain audio frame [34].

where the masking ability of tone-masking-noise and noise-masking-tone is considered, by exploiting the tonality index in each partition.

The masking threshold is transformed back to the linear frequency scale by spreading it evenly over all spectral lines corresponding to the partitions, as seen in Figure 11.10 in preparation for the calculation of the Signal-to-Mask Ratios (SMR) for each subband. The minimum masking threshold, as shown in Figure 11.10, takes into account the value of the threshold in quiet, T_q, raising the masking threshold value to the value of T_q, if the masking threshold value is lower than T_q. Finally, the SMR is computed for each scalefactor band as the ratio of the signal energy within a frequency-domain scalefactor band to the minimum masking threshold for that particular band, as depicted graphically in Figure 11.10. The SMR values will then be used for the subsequent allocation of bits in each frequency band.

11.2.4 Temporal Noise Shaping

Temporal Noise Shaping (TNS) in audio coding was first introduced by Herre *et al.* in [451]. The TNS tool seen in Figure 11.7 is a frequency domain technique, which operates on the spectral coefficients generated by the analysis filterbank. The idea is to employ linear predic-

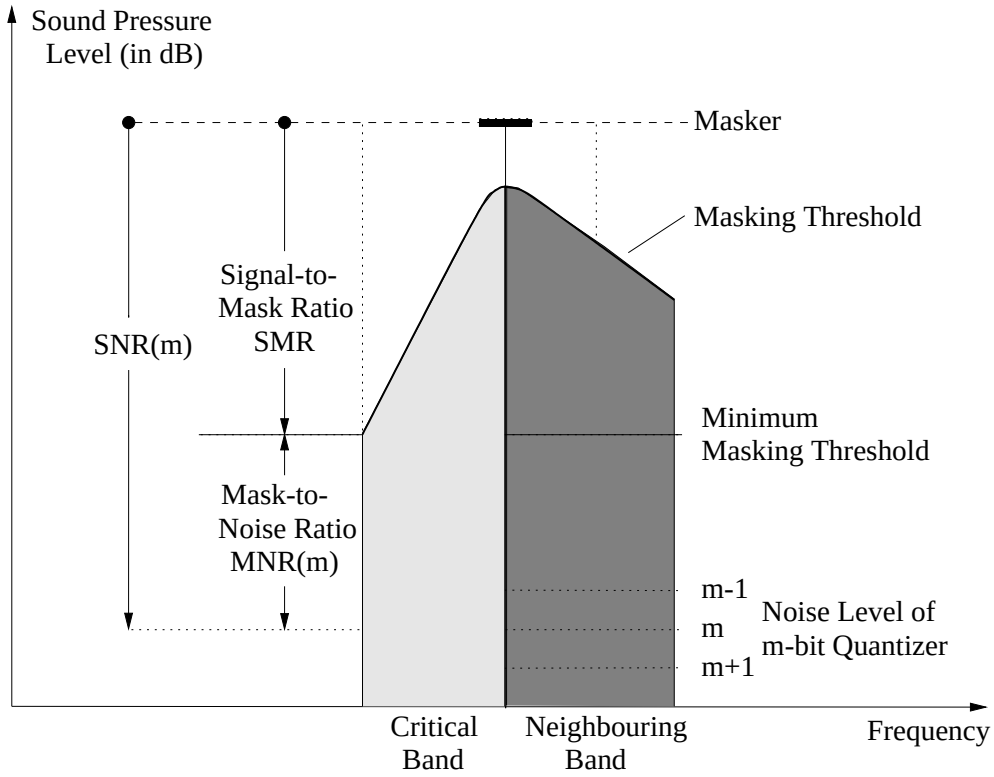

Figure 11.10: Masking effects and masking threshold calculation.

tive coding across the frequency range, rather than in the time-domain. TNS is particularly important, when coding signals that vary dynamically over time, such as for example transient signals. Transform codecs often encounter problems when coding such signals since the distribution of the quantization noise can be controlled over the frequency range but this spectral noise shaping is typically time-invariant over a complete transform block. When a signal changes drastically within a time-domain transform block without activating a switch to shorter time-domain transform lengths, the associated time-invariant distribution of quantization noise may lead to audible audio artifacts.

The concept of TNS is based upon the time- and frequency-domain duality of the LPC analysis paradigm [433], since it is widely recognized that signals exhibiting a non-uniform spectrum can be efficiently coded either by directly encoding the spectral-domain transform coefficients using transform coding, or by applying linear predictive coding methods to the time-domain input signal. The corresponding 'duality statement' relates to the encoding of audio signals exhibiting a time-variant time-domain behaviour, such as in case of transient signals. Thus, efficient encoding of transient signals can be achieved by either directly encoding their time domain representation or by employing predictive audio coding methods across the frequency domain.

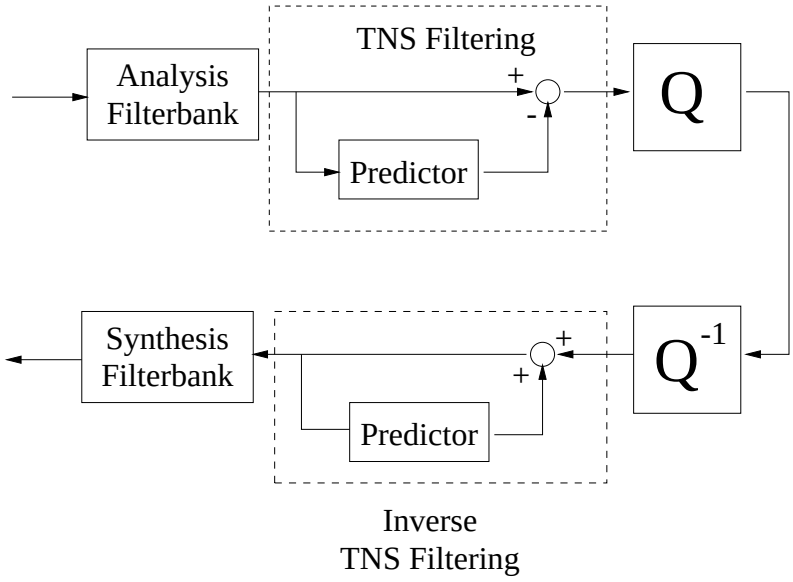

Figure 11.11: The TNS processing block also seen in Figure 11.7.

Figure 11.11 shows the more detailed TNS filtering process seen in the centre of Figure 11.7. The TNS tool is applied to the spectral-domain transform coefficients after the filterbank stage of Figure 11.7. The TNS filtering operation replaces the spectral-domain coefficients with the prediction residual between the actual and predicted coefficient values, thereby increasing their representation accuracy. Similarly, at the decoder an inverse TNS filtering operation is performed on the transform coefficient prediction residual in order to obtain the decoded spectral coefficients. TNS can be applied to either the entire frequency spectrum, or only to a part of the spectrum, such that the frequency-domain quantization can be controlled in a time-variant fashion [40], again, with the objective of achieving agile and responsive adjustment of the frequency-domain quantization scheme for sudden time-domain transients. In combination with further techniques such as window switching and gain control, the pre-echo problem can be further mitigated. In addition, the TNS technique enables the peak bit rate demand of encoding transient signals to be reduced. Effectively, this implies that an encoder may stay longer in the conventional and more bit-rate efficient long encoding block.

Additionally, the long-term time-domain redundancy of the input signal may be exploited using the well-documented Long Term Prediction (LTP) technique, which is frequently used in speech coding [41, 335, 367].

11.2.5 Stereophonic Coding

The MPEG-4 scheme includes two specific techniques for encoding stereo coding of signals, namely intensity-based stereo coding [452] and Mid/Side (M/S) stereo coding [453], both

of which will be described in this section. These coding strategies can be combined by selectively applying them to different frequency regions.

Intensity-based stereophonic coding is based on the analysis of high-frequency audio perception, as outlined by Herre *et al.* in [452]. Specifically, high-frequency audio perception is mainly based on the energy-time envelope of this region of the audio spectrum. It allows a stereophonic channel pair to share a single set of spectral intensity values for the high-frequency components with little or no loss in sound quality. Effectively, the intensity signal spectral components are used to replace the corresponding left channel spectral coefficients, while the corresponding spectral coefficients of the right channel are set to zero. Intensity-based stereophonic coding can also be interpreted as a simplified approximation to the idea of directional coding. Thus, only the information of one of the two stereo channel is retained, while the directional information is obtained with the aid of two scalefactor values assigned to the left and right channels [454].

On the other hand, M/S stereo coding allows the pair of stereo channels to be conveyed as left/right (L/R) or as the mid/side (M/S) signals representing the M/S information on a block-by-block basis [453], where M=(L+R)/2 and S=(L-R)/2. Here, The M/S matrix takes the sum information M + S, and sends it to the left channel, and the difference information M - S, and sends it to the right channel. When the left and right signals are combined, (M + S) + (M - S) = 2M, the sum is M information only. The number of bit actually required to encode the M/S information and L/R information is then calculated. In cases where the M/S channel pair can be represented with the aid of fewer bits, while maintaining a certain maximum level of quantization distortion, the corresponding spectral coefficients are encoded, and a flag bit is set for signalling that the block has utilized M/S stereo coding. During decoding the decoded M/S channel pair is converted back to its original left/right format.

11.2.6 AAC Quantization and Coding

After all the pre-processing stages of Figure 11.7 using various coding tools, as explained in earlier sections, all parameters to be transmitted will now have to be quantized. The quantization procedure follows an analysis-by-synthesis process, consisting of two nested iteration loops, which are depicted in Figure 11.12. This involves the non-uniform quantization of the spectral-domain transform coefficients [40]. Transform-domain non-linear quantizers have the inherent advantage of facilitating spectral-domain noise shaping in comparison to conventional linear quantizers [431]. The quantized spectral-domain transform coefficients are then coded using Huffman coding. In order to improve the achievable subjective audio quality, the quantization noise is further shaped using scalefactors [455], as it is highlighted below.

Specifically, the spectrum is divided into several groups of spectral-domain transform coefficients, which are referred to as scalefactor bands (SFB). Each frequency-domain scalefactor band will have its individual scalefactor, which is used to scale the amplitude of all spectral-domain transform coefficients in that scalefactor band. This process shapes the spectrum of the quantization noise according to the masking threshold portrayed in Figure 11.10, as estimated on the basis of the psychoacoustic model. The width of the frequency-domain scalefactor bands is adjusted according to the critical bands of the human auditory system [423], seen in Figure 11.9. The number of frequency-domain scalefactor bands and their width depend on the transform length and sampling frequency. The spectral-domain noise shaping is achieved

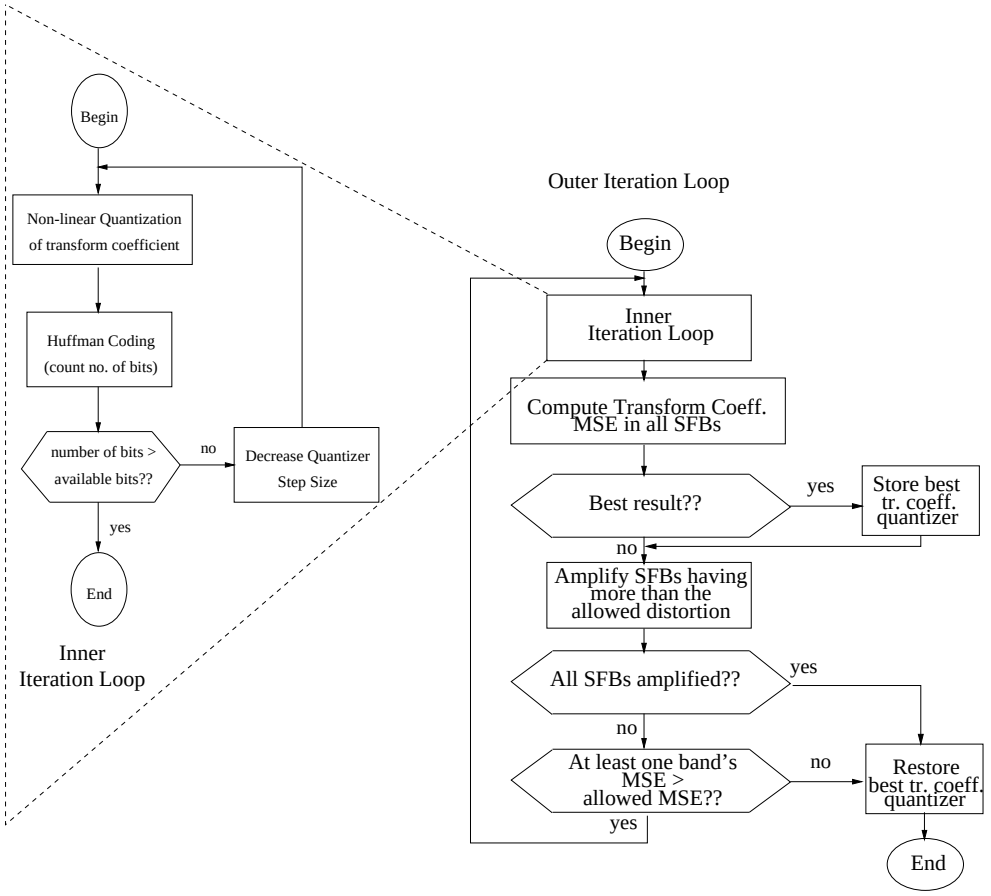

Figure 11.12: AAC inner and outer quantization loops designed for encoding the frequency-domain transform coefficients.

by adjusting the scalefactor using a step size of 1.5 dB. The decision as to which scalefactor bands should be amplified/attenuated relies on the threshold computed from the psychoacoustic model and also on the number of bits available. The spectral coefficients amplified have high amplitudes and this results in a higher SNR after quantization in the corresponding scalefactor bands. This also implies that more bits are needed for encoding the transform coefficients of the amplified scalefactor bands and hence the distribution of bits across the scalefactor bands will be altered. Naturally, the scalefactor information will be needed at the decoder, hence the scalefactors will have to be encoded as efficiently as possible. This is achieved by first exploiting the fact that the scalefactors usually do not change dramatically from one scalefactor band to another. Thus a differential encoding proved useful. Secondly, Huffman coding is applied, in order to further reduce the redundancy associated with the encoding of the scalefactors [40].

Again, the AAC quantization and coding process consists of two iteration loops, the in-

ner and outer loops. The inner iteration loop shown in Figure 11.12 consists of a non-linear frequency-domain transform coefficient quantizer and the noiseless Huffman coding module. The frequency-domain transform coefficient values are first quantized using a non-uniform quantizer, and further processing using the noiseless Huffman coding tool is applied for achieving a high coding efficiency. The quantizer step size is decreased until the number of bits generated exceeds the available bit rate budget of the particular scalefactor band considered. Once the inner iteration process is completed, the outer loop evaluates the Mean Square Error (MSE) associated with all transform coefficients for all scalefactor bands. The task of the outer iteration loop is to amplify the transform coefficients of the scalefactor bands, in order to satisfy the requirements of the psychoacoustic model. The MSE computed is compared to the masking threshold value obtained from the associated psychoacoustic analysis. When the best result, i.e. the lowest MSE is achieved, the corresponding quantization scheme will be stored in memory. Subsequently, the scalefactor bands having a higher MSE than the acceptable threshold are amplified, using a step size of 1.5 dB. The iteration process will be curtailed, when all scalefactor bands have been amplified or it was found that the MSE of no scalefactor band exceeds the permitted threshold. Otherwise, the whole process will be repeated, using new SFB amplification values, as seen in Figure 11.12.

11.2.7 Noiseless Huffman Coding

The noiseless Huffman coding tool of Figure 11.12 is used for further reducing the redundancy inherent in the quantized frequency-domain transform coefficients of the audio signal. One frequency-domain transform coefficients quantizer per scalefactor band is used. The step size of each of these frequency-domain transform coefficients quantizers is specified in conjunction with a global gain factor that normalizes the individual scalefactors. The global gain factor is coded as an 8-bit unsigned integer. The first scalefactor associated with the quantized spectrum is differentially encoded relative to the global gain value and then Huffman coded using the scalefactor codebook. The remaining scalefactors are differentially encoded relative to the previous scalefactor and then Huffman coded using the scalefactor codebook.

Noiseless coding of the quantized spectrum relies on partitioning of the spectral coefficients into sets. The first partitioning divides the spectrum into scalefactor bands that contain an integer multiple of 4 quantized spectral coefficients. The second partitioning divides the quantized frequency-domain transform coefficients into sections constituted by several scalefactor bands. The quantized spectrum within such a section will be represented using a single Huffman codebook chosen from a set of twelve possible codebooks. This includes a particular codebook that is used for signalling that all the coefficients within that section are zero. Hence no spectral coefficients or scalefactors will be transmitted for that particular band, and thus an increased compression ratio is achieved. This is a dynamic quantization process, which varies from block to block, such that the number of bits needed for representing the full set of quantized spectral coefficients is minimized. The bandwidth of the section and it AUTOINDEX number=27 closes associated Huffman codebook indices must be transmitted as side information, in addition to the section's Huffman coded spectrum.

Huffman coding creates variable length codes [431,456], where higher probability symbols are encoded by shorter codes. The Huffman coding principles are highlighted in Figure 11.13. Specifically, successive Column 0 in Figure 11.13 shows the set of symbols A, B, C and D, which are Huffman coded in the successive columns. At first, the symbols are sorted from

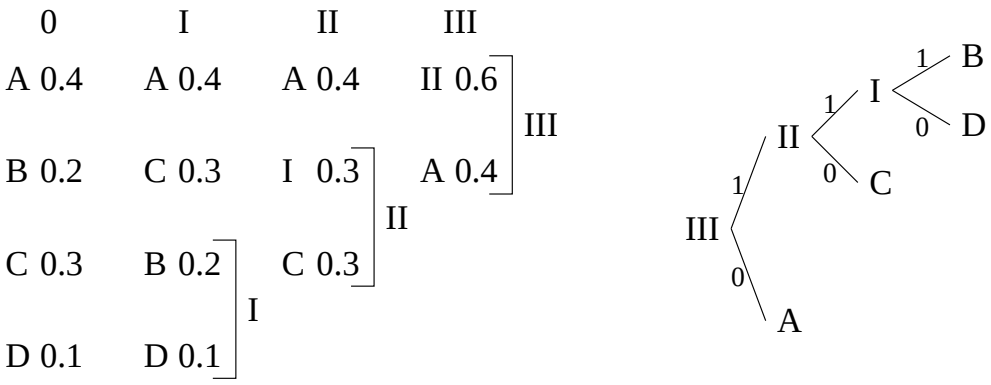

Figure 11.13: Huffman coding

top to bottom with decreasing probability. In every following step, the two lowest probability symbols at the bottom are combined to one symbol, which is assigned the sum of the single probabilities. The new symbol is then fitted into the list at the correct position according to its new probability of occurrence. This procedure is continued, until all codewords are merged, which leads to a coding tree structure, as seen in Figure 11.13. The assignment of Huffman coded bits is carried out as follows. At every node, the upper branch is associated with a binary '1', and the lower branch with a binary '0', or the other way round. The complete binary tree can be generated by recursively reading out the symbol list, starting with symbol 'III'. As a result, symbol A is coded as '0', B with '1111', C as '10' and D with '110'since none of the symbols constitutes a prefix of the other symbols, their decoding is unambiguous.

11.2.8 Bit-Sliced Arithmetic Coding

The Bit-Sliced Arithmetic Coding (BSAC) tool, advocated by Park *et al.* [449] is an alternative to the AAC noiseless Huffman coding module of Section 11.2.7, while all other modules of the AAC-based codec remain unchanged, as shown earlier in Figure 11.7. BSAC is included in the MPEG-4 Audio Version 2 for supporting finely-grained bitstream scalability, and further reducing the redundancy inherent in the scalefactors and in the quantized spectrum of the MPEG-4 T/F codec [457].

In MPEG-4 Audio Version 1, the General Audio (GA) codec supports coarse scalability where a base layer bitstream can be combined with one or more enhancement layer bitstreams in order to achieve a higher bit rate and thus an improved audio quality. For example, in a typical scenario we may utilise a 24 kbit/s base layer together with two 16 kbit/s enhancement layers. This gives us the flexibility of decoding in three modes, namely 24 kbit/s, 24+16=40 kbit/s or 24+16+16=56 kbit/s modes. Each layer carries significant amount of side information and hence finely-grained scalability was not supported efficiently in Version 1.

The BSAC tool provides scalability in steps of 1 kbit/s per channel. In order to achieve finely-grained scalability, a 'bit-slicing' scheme is applied to the quantized spectral coefficients [449]. A simple illustration assisting us in understanding the operation of this BSAC algorithm is shown in Figure 11.14. Let us consider a quantized transform coefficient se-

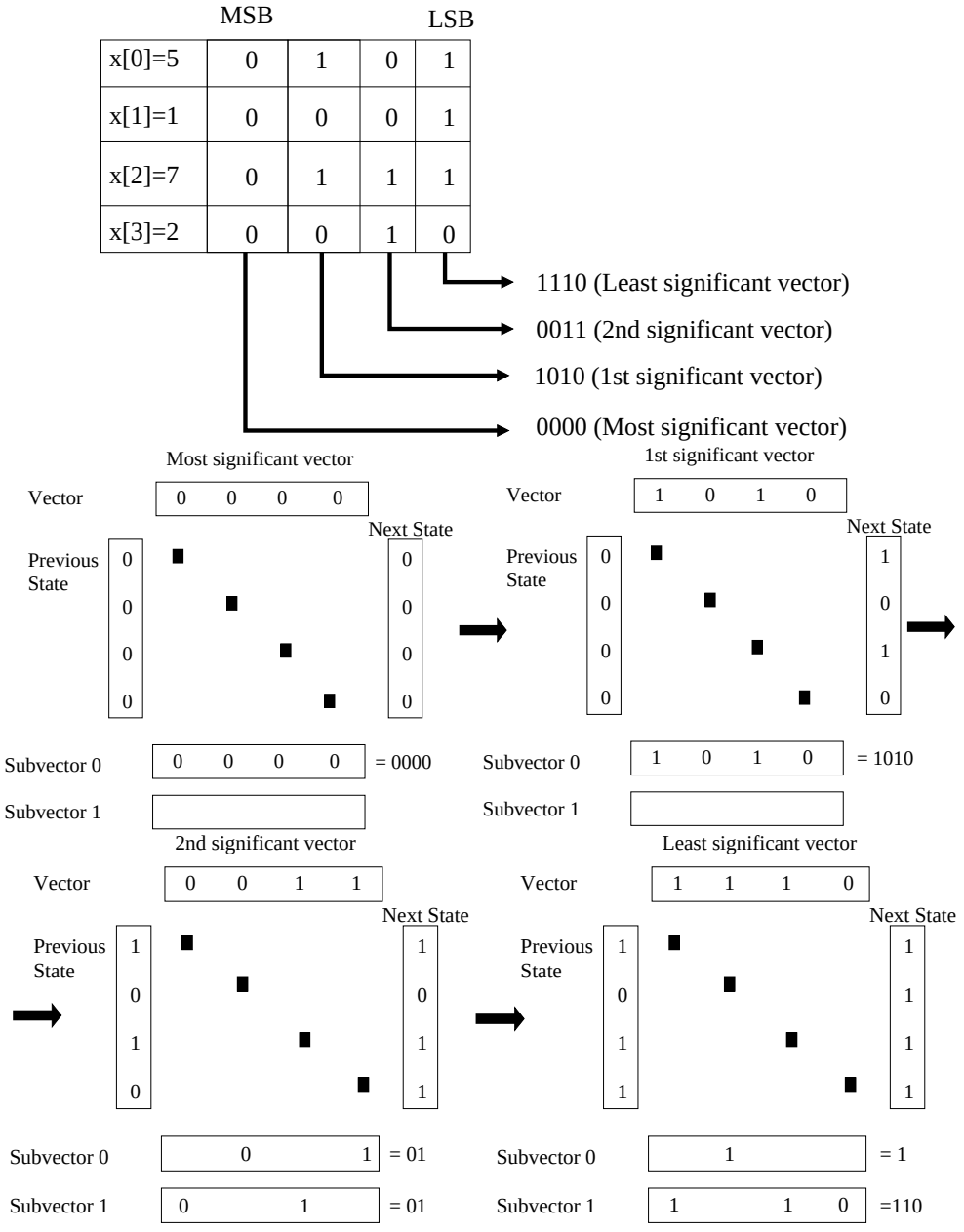

Figure 11.14: BSAC bit-sliced operations, where four quantized bit-sliced sequence are mapped into four 4-bit vectors.

quence, x[n], each coefficient quantized with the aid of four bits, assuming the values of x[0]=5, x[1]=1, x[2]=7 and x[3]=2. Firstly, the bits of this group of sequences are processed in slices according to their significance, commencing with the MSB or LSB. Thus, the Most Significant Bits (MSB) of the quantized vectors are grouped together yielding the bit-sliced vector of 0000, followed by the 1st significant vector (1010), 2nd significant vector (0011) and the least significant vector (1110), as displayed in the top half of Figure 11.14.

The next step is to process the four bit-sliced vectors, exploiting their previous values, which is first initialized to zero. The MSB vector (0000) is first decomposed into two subvectors. Subvector 0 is is composed of the bit values of the current vector whose previous state is 0, while Subvector 1 consists of bit values of the current vector whose previous state is 1. Note that when a specific previous state bit is zero, the next state bit will remain zero if the corresponding bit value of the current vector is zero and it is set to 1, when either the previous state bit or the current vector's bit value, or both is 1.

By utilising this BSAC scheme, finely-grained bit rate scalability can be achieved by employing first the most significant bits. An increasing number of enhancement layers can be utilised by using more of the less significant bits obtained through the bit-slicing procedure. The actively encoded bandwidth can also be increased by providing bit slices of the transform coefficients in the higher frequency bands.

11.2.9 Transform-domain Weighted Interleaved Vector Quantization

As shown in Figure 11.7, the third quantization and coding tool employed for compressing the spectral components is the so-called Transform-domain Weighted Interleaved Vector Quantization (TWINVQ) [41] scheme. It is based on an interleaved vector quantization and LPC spectral estimation technique, and its performance was superior in comparison to AAC coding at bit rates below 32 kbit/s per channel [450, 458–460]. TWINVQ invokes some of the compression tools employed by the G.729 8 kbit/s standard codec [372], such as the LPC analysis, LSF parameter quantization employing conjugate structure VQ [461]. The operation of the TWINVQ encoder is shown in Figure 11.15. Each block will be described during our further discourse in a little more depth. Suffice to say that TWINVQ was found to be superior for encoding audio signals at extremely low bit rates, since the AAC codec performs poorly at low bit rates, while the CELP mode of MPEG-4 is unable to encode music signals [462]. The TWIN-VQ scheme has also been used as a general coding paradigm for representing both speech and music signals at a rate of 1 bit per sample [463].

More specifically, the input signal, as shown in Figure 11.15, is first transformed into the frequency domain using the MDCT. Before the transformation, the input signal is classified into one of three modes, each associated with a different transform window size, namely a long, medium or short window. In the long-frame mode, the transform size is equal to the frame size of 1024. The transform operations are carried out twice in a 1024-sample frame with a half-transform size in the medium-frame mode, and eight times having a one-eighth transform size in the short-frame mode. These different window sizes cater for different input signal characteristics. For example, transient signals are best encoded using a small transform size, while stationary signals can be windowed employing the normal long frame mode.

As shown in Figure 11.15, the spectral envelope of the MDCT coefficients is approximated with the aid of LPC analysis applied to the time-domain signal. The LPC coefficients are then transformed to the Line Spectrum Pair (LSP) parameters. A two-stage split vector quantizer

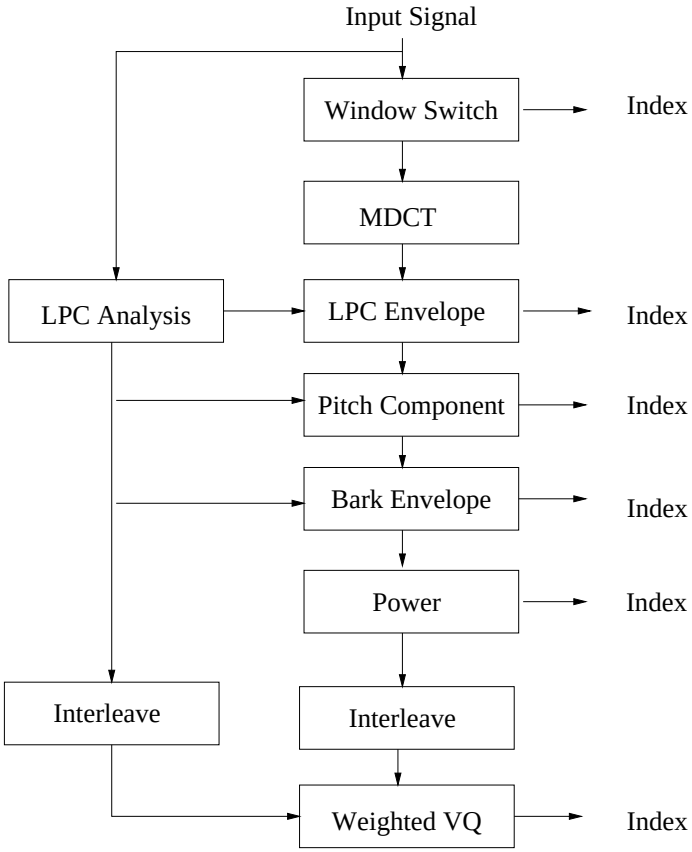

Figure 11.15: TWINVQ encoder [458].

with inter-frame moving-average prediction was used for quantizing the LSPs, which was also employed in the G.729 8 kbit/s standard codec [372]. The MDCT coefficients are then smoothed in the frequency domain using this LPC spectral envelope. After the smoothing by the LPC envelope, the resultant MDCT coefficients still retain their spectral fine structure. In this case, the MDCT coefficients would still exhibit a high dynamic range, which is not amenable to vector quantization. Pitch analysis is also employed, in order to obtain the basic harmonic of the MDCT coefficients, although this is only applied in the long frame mode. The periodic MDCT peak components correspond to the pitch period of speech or audio signal. The extracted pitch parameters are quantized by the interleaved weighted vector quantization scheme [464], as it will be explained later in this section.

As seen in Figure 11.15, the Bark-envelope is then determined from the MDCT coefficients, which is smoothed by the LPC spectrum. This is achieved by first calculating the square-rooted power of the smoothed MDCT coefficients corresponding to each Bark-scale subband. Subsequently, the average MDCT coefficient magnitudes of the Bark-scale subbands are normalized by their overall average value in order to create the Bark-scale enve-

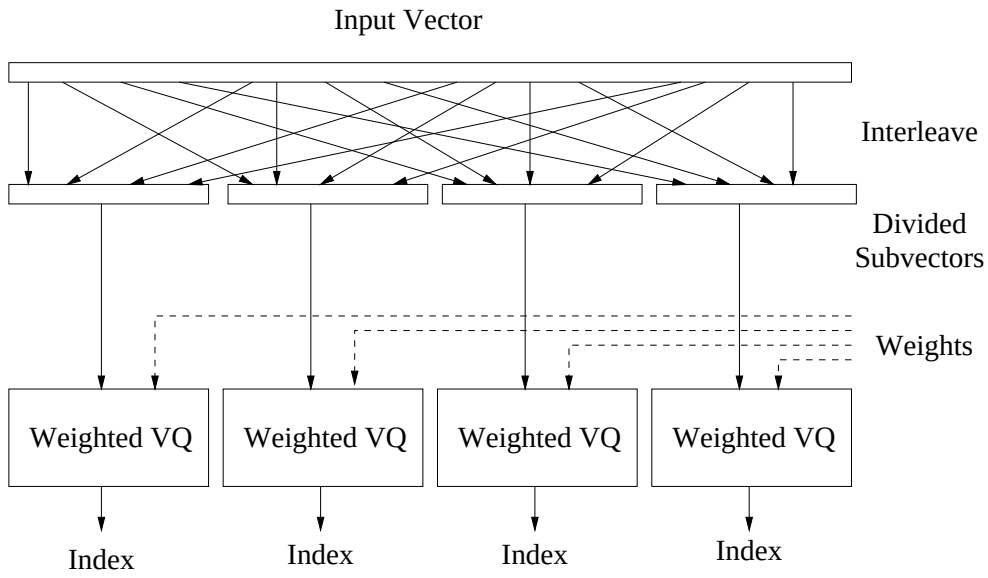

Figure 11.16: TWINVQ interleaved weighted vector quantization process [41].

lope. Before quantizing the Bark-scale envelope, further redundancy reduction is achieved by employing interframe backward prediction, whereby the correlation between the Bark-scale envelope of the current 23.2 ms frame and that of the previous frame is exploited. If the correlation is higher than 0.5, the prediction is activated. Hence, an extra flag bit has to be transmitted. The Bark-scale envelope is then vector quantized using the technique of interleaved weighted vector quantization, as seen at the bottom of Figure 11.15 [463] and augmented below.

At the final audio coding stage, the smoothed MDCT coefficients are normalized by a global frequency-domain gain value, which is then scalar quantized in the logarithm domain, which takes place in the 'Weighted VQ' block of Figure 11.15. Finally, the MDCT coefficients are interleaved, divided into subvectors for the sake of reducing the associated matching complexity, and vector quantized using a weighted distortion measure derived from the LPC spectral envelope [464]. The role of the weighting is that of reducing the spectral-domain quantization errors in the perceptually most vulnerable frequency regions. Moriya *et al.* [464] proposed this vector quantizer, since it constitutes a promising way of reducing the computational complexity incurred by vector quantization [461], as it will be highlighted below. Specifically, this two-stage MDCT VQ-scheme uses two sets of trained codebooks for vector quantizing the MDCT coefficients of a subvector, and the MDCT subvector is reconstructed by superimposing the two codebook vectors. In the encoder, a full-search is invoked for finding the combination of the code vector indices that minimizes the distortion between the input and reconstructed MDCT subvector. This two-stage MDCT VQ-scheme constitutes a sub-optimal arrangement in comparison to a single-stage VQ, however it significantly reduces the memory and the computational complexity required. The employment of a fixed frame rate combined with the above vector quantizer improves its robustness against

Parameters	No. of Bits
Window mode	4
MDCT coefficients	295
Bark-envelope VQ	44
Prediction switch	1
Gain factor	9
LSF VQ	19
Total bits	372

Table 11.1: MPEG-4 TWINVQ bit allocation scheme designed for a rate of 16 kbit/s, which corresponds to 372 bits per 23.22 ms frame.

errors, since it does not use any error sensitive compression techniques, such as adaptive bit allocation or variable length codes [458].

The MPEG-4 TWINVQ bitstream structure is shown in Table 11.1 in its 16 kbit/s mode, which will be used in our investigations in order to construct a multi-mode speech transceiver, as detailed in Section 11.5. A substantial fraction of the bits were allocated for encoding the MDCT coefficients, which were smoothed by the LPC and Bark-scale spectra. Specifically, a total of 44 bits were allocated for vector quantizing the Bark-scale envelope, while one bit is used for the interframe prediction flag. Nine bits were used for encoding the global spectral-domain gain value obtained from the MDCT coefficients and the LSF VQ requires 19 bits per 23.22 ms audio frame.

11.2.10 Parametric Audio Coding

An enhanced functionality provided by the MPEG-4 Audio Version 2 scheme is parametric audio coding, with substantial contributions from Purnhagen *et al.* [465–467], Edler [468], Levine [469] and Verma [470]. This compression tool facilitates the encoding of audio signals at the very low bit rate of 4 kbit/s, using a parametric representation of the audio signal. Similarly to the philosophy of parametric speech coding, here instead of waveform coding the audio signal is decomposed into audio objects, which are described by appropriate source models and the quantized models parameters are transmitted. This coding scheme is referred to as the Harmonic and Individual Lines plus Noise (HILN) technique, which includes object models for sinusoids, harmonic tones and noise components [466].

Due to the limited bit rate budget at low target bit rate, only the specific parameters that are most important for maintaining an adequate perceptual quality of the signal are transmitted. More specifically, in the context of the HILN technique, the frequency and amplitude parameters are quantized using existing masking rules from psychoacoustics [424]. The spectral envelope of the noise and harmonic tones is described using LPC techniques. Parameter prediction is employed in order to exploit the correlation between the parameters across consecutive 23.22 ms frames. The quantized parameters are finally encoded using high-efficiency, but error-sensitive Huffman coding. Using a speech/music classification tool in the encoder, it is possible to automatically activate the coding of speech signals using the HVXC parametric encoder or the HILN encoder contrived for music signals.

The operating bit rate of the HILN scheme is at a fixed rate of 6 kbit/s in the mono, 8kHz

sampling rate mode and 16 kbit/s in the mono, 16 kHz sampling rate mode, respectively. In an alternative proposal by Levine *et al.* in [471], an audio codec employing switching between parametric and transform coding based representations was advocated. Sinusoidal signals and noise are modelled using multiresolution sinusoidal modelling [469] and Bark-scale based noise modelling, respectively, while the transients are represented by short-window based of transform coding. Verma *et al.* in [470] extended the work in [469] by proposing an explicit transient model for sinusoidal-like signals and for noise. A slowly varying sinusoidal signal is impulse-like in the frequency domain. By contrast, transients are impulse-like in the time domain and cannot be readily represented with the aid of Short-Time Fourier Transform (STFT) based analysis. However, due to the duality between time and frequency, transients which are impulse-like in the time domain, appear to be oscillatory in the frequency domain. Hence, sinusoidal modelling can be applied after the transformation of the transient time-domain signals to sinusoidal-like signals in the frequency domain by quantizing their DCT [470] coefficients.

11.3 Speech Coding in MPEG-4 Audio

While the employment of transform coding is dominant in coding music audio and speech signals at rates above 24 kbit/s, its performance deteriorates, as the bit rate decreases. Hence, in the MPEG-4 audio scheme, dedicated speech coding tools are included, operating at the bit rates in the range between 2 and 24 kbit/s [44, 49]. Variants of the Code Excite Linear Prediction (CELP) technique [365] are used for the encoding of speech signals at the bit rates between 4 and 24 kbit/s, incorporating the additional flexibility of encoding speech represented at both 8 and 16 kHz sampling rates. Below 4 kbit/s, a sinusoidal technique, namely the so-called Harmonic Vector eXcitation Coding (HVXC) scheme was selected for encoding speech signals at rates down to a bit rate of 2 kbit/s. The HVXC technique will be described in the next section, while CELP schemes will be discussed in Section 11.3.2.

11.3.1 Harmonic Vector Excitation Coding

Harmonic Vector Excitation Coding (HVXC) is based on the signal classification of voiced and unvoiced speech segments, facilitating the encoding of speech signals at 2 kbit/s and 4 kbit/s [472, 473]. Additionally, it also supports variable rate encoding by including specific coding modes for both background noise and mixed voice generation in order to achieve an average bit rate as low as 1.2 - 1.7 kbit/s.

The basic structure of an HVXC encoder is shown in Figure 11.17, which first performs LPC analysis for obtaining the LPC coefficients. The LPC coefficients are then quantized and used in the inverse LPC filtering block in order to obtain the prediction residual signal. The prediction residual signal is then transformed into the frequency domain using the Discrete Fourier Transform (DFT) and pitch analysis is invoked, in order to assist in the V/UV classification process. Furthermore, the frequency-domain spectral envelope of the prediction residual is quantized by using a combination of two-stage shape vector quantizer and a scalar gain quantizer. For unvoiced segments, a closed-loop codebook search is carried out in order to find the best excitation vector.

Specifically, the HVXC codec operates on the basis of a 20 ms frame length for speech

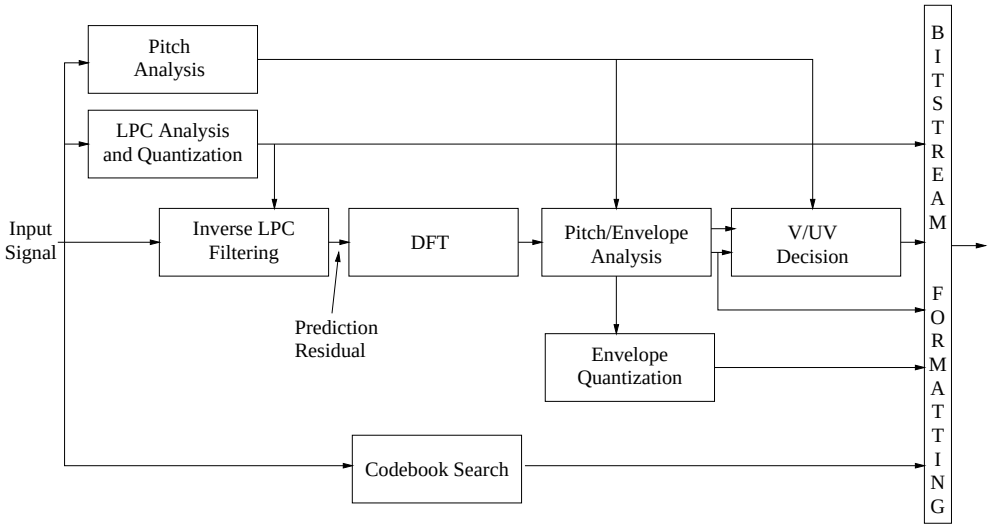

Figure 11.17: Harmonic Vector Excitation Coding

signals represented at an 8 kHz sampling rate. Table 11.2 shows the bit allocation schemes of the HVXC codec at rates of 2 and 4 kbit/s [474]. Both voiced and unvoiced speech segments use the LSF parameters and the voiced/unvoiced indicator flag. For 2 kbit/s transmission of voiced speech, the parameters include 18 bits for LSF quantization using a two-stage split vector quantizer, which facilitates the reduction of the codebook search complexity by mitigating the VQ matching complexity. Furthermore, 2 bits are used for the V/UV mode indication, where the extra one bit is used to indicate the background noise interval and mixed speech modes for variable rate coding, as will be explained later. Furthermore, 7 bits are dedicated to pitch encoding, while 8 and 5 bits are used for encoding the harmonic shape and gain of the prediction residual in Figure 11.17, respectively. Explicitly, for the quantization of the harmonic spectral magnitudes/shapes of the prediction residual in Figure 11.17, a two-stage shape vector quantizer is used, where the size of both the shape codebooks is 16, both requiring a four-bit index. The codebook gains are quantized using three and two bits, respectively. In the case of unvoiced speech transmission at 2 kbit/s, besides the LSF quantization indices and the V/UV indication bits, the shape and gain codebook indices of the Vector eXCitation (VXC) requires 6 and 4 bits, respectively for a 10 ms frame length.

For 4 kbit/s transmission, a coding enhancement layer is added to the base layer of 2 kbit/s. In the case of LSF quantization, a 10-dimensional vector quantizer using an 8-bit codebook is added to the 18 bits/20 ms LSF quantizer scheme of the 2 kbit/s codec mode seen at the top of Table 11.2. This results in an increased bit rate requirement for LSF quantization, namely from 18 bits/20ms to 26 bits/20ms. A split VQ scheme, composed of four vector quantizers having addresses of 7, 10, 9 and 6 bits, respectively is added to the two-stage vector quantizer required for the quantization of the harmonic shapes of the prediction residual in Figure 11.17. This results in a total of bit rate budget increase of 32 bits/20 ms, as seen in Table 11.2. For unvoiced speech segment encoding at 4 kbit/s, the excitation vectors of the

	Voiced	Common	Unvoiced
LSF1 (2-stage Split VQ at 2 kbit/s		18bits/20ms	
LSF2 (at 4 kbit/s)		8bits/20ms	
V/UV		2bits/20ms	
Pitch	7bits/20ms		
Harmonic1 Shape (at 2 kbit/s)	4+4bits/20ms		
Harmonic1 Gain (at 2 kbit/s)	5bits/20ms		
Harmonic2 Split (at 4 kbit/s)	32bits/20ms		
VXC1 Shape (at 2 kbit/s)			6bits/10ms
VXV1 Gain (at 2 kbit/s)			4bits/10ms
VXC2 Shape (at 4 kbit/s)			5bits/5ms
VXC2 Gain (at 4 kbit/s)			3bits/5ms
2 kbit/s mode	40bits/20ms		40bits/20ms
4 kbit/s mode	80bits/20ms		80bits/20ms

Table 11.2: MPEG-4 Bit allocations at the fixed rates of 2.0 and 4.0 kbit/s using the HVXC coding mode [41].

Mode	Background Noise	Unvoiced	Mixed Voiced/Voiced
V/UV	2bits/20ms	2bits/20ms	2bits/20ms
LSF	0bits/20ms	18bits/20ms	18bits/20ms
Excitation	0bits/20ms	8bits/20ms (gain only)	20bits/20ms (pitch & harmonic spectral parameters)
Total	2bits/20ms = 0.1 kbit/s	28bits/20ms = 1.4 kbit/s	40bits/20ms = 2.0kbit/s

Table 11.3: Bit allocations for variable rate HVXC coding [41].

enhancement layer are obtained by utilising codebook search and the gain/shape codebook indices, which minimize the weighted distortion are transmitted. Specifically, a 5-bit shape codebook as well as 3-bit gain codebook are used and this procedure is updated every 5 ms. For the unvoiced speech segments, the LPC coefficients of only the current 20 ms frame are used for two 10 ms subframes without any interpolation procedure using the LPC coefficients from the previous frame. Again, the codec's performance was summarised in Table 11.2.

Optional variable rate coding can be applied to the HVXC codec, incorporating background noise detection, where only the mode bits are received during the "background noise mode". When the HVXC codec is in the "background noise mode", the decoding is similar to the manner applied in an UV frame, but in this scenario no LSF parameters are transmitted while only the mode bits are transmitted. Instead two sets of LSF parameters generated during the previous two UV frames will be used for the LPC synthesis process. During the background noise mode, fully encoded unvoiced (UV) frames are inserted every nine 20 ms frames, in order to transmit the background noise parameters. This means only eight consecutive "background noise" frame are allowed to use the same two sets of LSF parameters

CELP		
Mode	Narrowband	Wideband
Sampling Rate (kHz)	8	16
Bandwidth (Hz)	300 - 3400	50 - 7000
Bit Rate (kbit/s)	3.85 - 12.2	10.9 - 24.0
Excitation Scheme	MPE/RPE	RPE
Frame Size (ms)	10 - 40	10 - 20
Delay (ms)	15 - 85	18.75 - 41.75
Features	Multi Bit Rate Coding Bit Rate Scalability Bandwidth Scalability Complexity Scalability	

Table 11.4: Summary of various features of the MPEG-4 CELP codec [41].

from the previous UV frames. Hereafter new UV frame will be transmitted. This UV frame may or may not be a real UV frame indicating the beginning of active speech bursts. This is signalled by the transmitted gain factor. If the gain factor is smaller or equal to the previous two gain values, then this UV frame is regarded as background noise. In this case the most recent previously transmitted LSF parameters are used for maintaining the smooth variation of the LSF parameters. Otherwise, the currently transmitted LSFs are used, since the frame is deemed a real UV frame. During background noise periods, a gain-normalised Gaussian noise vector is used instead of the stochastic prediction residual shape codebook entry employed during UV frame decoding. The prediction residual gain value is encoded using an 8-bit codebook entry, as displayed in Table 11.3.

Table 11.3 shows the bit allocation scheme of variable rate HVXC coding for four different encoding modes, which are the modes dedicated to background noise, unvoiced, mixed voiced and voiced segments. The mixed voiced and voiced modes share the same bit allocation at 2 kbit/s. The unvoiced mode operates at 1.4 kbit/s, where only the gain parameter of the vector excitation is transmitted. Finally, for the background noise mode only the two voiced/unvoiced/noise signalling bits are transmitted.

11.3.2 CELP Coding in MPEG-4

While the HVXC mode of MPEG-4 supports the very low bit rate encoding of speech signals for rates below 4 kbit/s, the CELP compression tool is used for bit rates in excess of 4 kbit/s, as illustrated in the summary of Table 11.4. The MPEG-4 CELP tool enables the encoding of speech signals at two different sampling rates, namely at 8 and 16 kHz [475]. For narrowband speech coding, the operating bit rates are between 3.85 and 12.0 kbit/s. Higher bit rates between 10.9 and 24 kbit/s are allocated for wideband speech coding, which cater for a higher speech quality due to their extended bandwidth of about 7 kHz. The MPEG-4 CELP codec supports a range of further functionalities, which include the possibility of supporting multiple bit rates, bit rate scalability, bandwidth scalability and complexity scalability. Additionally, the MPEG-4 CELP mode supports both fixed and variable bit rate transmission. The

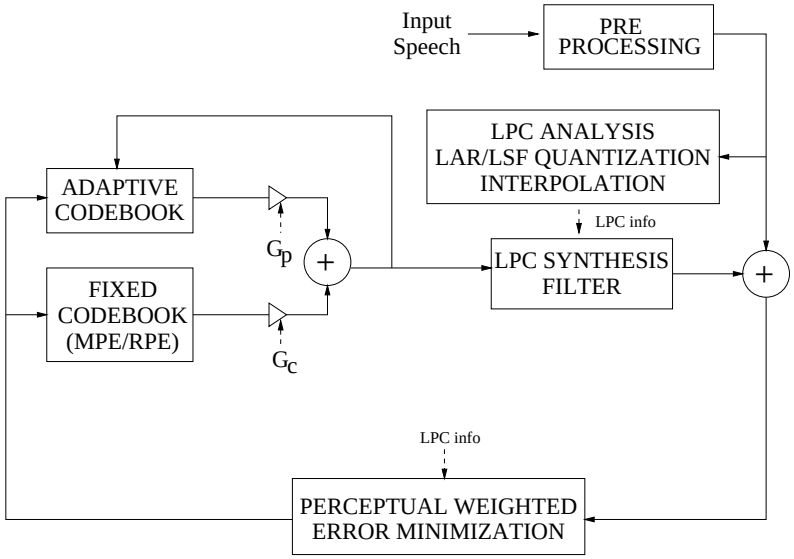

Figure 11.18: CELP encoder.

bit rate is specified by the user's requirements, taking account of the sampling rate chosen and also of the type of LPC quantizer (scalar quantizer or vector quantizer) selected. The default CELP codec operating at 16 kHz sampling rate employs a scalar quantizer and in this mode also the Fine Rate Control (FRC) switch is turned on. The FRC mode allows the codec to change the bit rate by skipping the transmission of the LPC coefficients, by utilising the **Interpolation** and the **LPC_Present** flags [476], as it will be discussed in Section 11.3.3. By contrast, at the 8 kHz sampling rate, the default MPEG-4 CELP mode utilises vector quantizer and the FRC switch is turned off.

As shown in Figure 11.18, first the LPC coefficients of the input speech are determined and converted to Log Area Ratios (LAR) or LSF. The LARs or LSFs are then quantized and also inverse quantized, in order to obtain the quantized LPC coefficients. These coefficients are used by the LPC synthesis filter. The excitation signal consists of the superposition of contributions by the adaptive codebook and one or more fixed codebooks. The adaptive codebook represents the periodic speech components, while the fixed codebooks are used for encoding the random speech components. The transmitted parameters include the LAR/LSF codebook indices, the pitch lag for the adaptive codebook, the shape codebook indices of the fixed codebook and the gain codebook indices of the adaptive as well as fixed codebook gains. Multi-Pulse Excitation (MPE) [477] or Regular Pulse Excitation (RPE) [478] can also be used for the fixed codebooks. The difference among the two lies in the degree of freedom for pulse positions. MPE allows more freedom in the choice of the inter-pulse distance than RPE, which has a fixed inter-pulse distance. As a result, MPE typically achieves a better speech coding quality than RPE at a given bit rate. On the other hand, the RPE scheme imposes a lower computational complexity than MPE, which renders MPE a useful tool for wideband speech coding, where the computational complexity is naturally higher than in

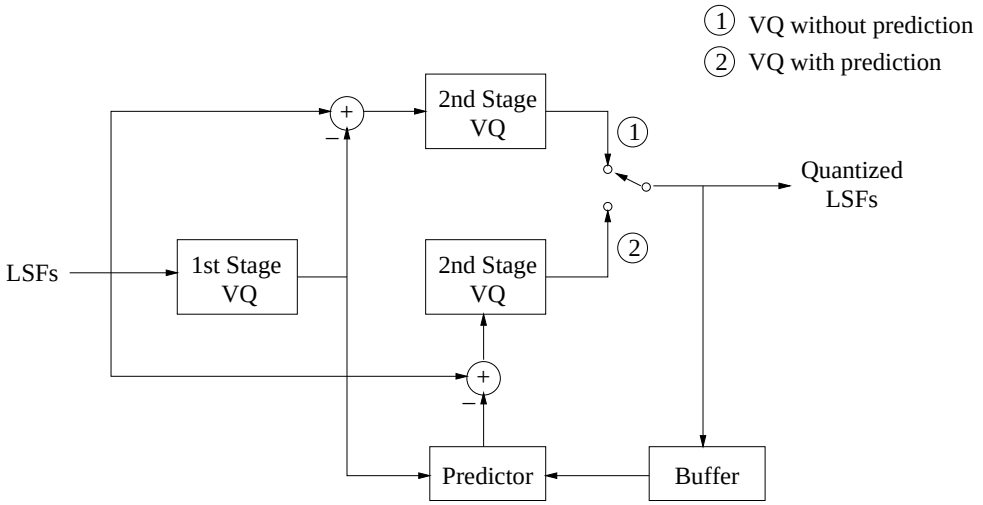

Figure 11.19: LSF VQ operating in two modes, with or without LSF prediction at the second stage VQ.

narrowband speech coding due to the doubled sampling rate used.

11.3.3 LPC Analysis and Quantization

Depending on the tolerable complexity, the LPC coefficients can be quantized using either a scalar or a vector quantization scheme. When a scalar quantizer is used, the LPC coefficients have to be transformed to the LAR parameters. In order to obtain the LAR parameters, the LPC coefficients are first transformed to the reflection coefficients [339]. The reflection coefficients are then quantized using a look-up table. The relationship between the LARs and the reflection coefficients are described by:

$$LAR[i] = log((1 - q\_rfc[i])/(1 + q\_rfc[i])) \tag{11.12}$$

where q_rfc represents the quantized reflection coefficients. The necessity to transmit the LARs depends on the amount of change between the current audio/speech spectrum and the spectrum described by the LARs obtained by interpolation from the LARs of the adjacent frame. If the spectral change is higher than a pre-determined threshold, then the current LAR coefficients are transmitted to the decoder. The threshold is adaptive, depending on the desired bit rate. If the resultant bit rate is higher than the desired bit rate, the threshold is raised, otherwise, it is lowered. In order to reduce the bit rate further, the LAR coefficients can be losslessly Huffman coded. We note however that lossless coding will only be applied to the LARs but not to the LSF, since only the LARs are scalar quantized and there is no LAR VQ in the standard.

If vector quantization of the LPC coefficients is used, the LPC coefficients are be transformed into the LSF domain. There are two methods of quantizing the LSFs in the CELP MPEG-4 mode. We can either employ a two-stage vector quantizer without interframe LSF

Interpolation	LPC_Present	Description
1	1	$LPC_{cur} = interpolate(LPC_{prev} + LPC_{next})$
0	0	$LPC_{cur} = LPC_{prev}$
0	1	LPC_{cur} = LPC received in current frame

Table 11.5: Fine Rate Control utilising the **Interpolation** and **LPC_Present** flags [41].

prediction, or in combination with interframe LSF prediction, as shown in Figure 11.19. In the case of using a two-stage vector quantizer without interframe LSF prediction, the second-stage VQ quantizes the LSF quantization error of the first stage. When interframe LSF prediction is employed, the difference between the input LSFs and the predicted LSFs is quantized. At the encoder, both methods are applied and the better method is selected by comparing the LSF quantization error, obtained by calculating the weighted mean squared LSF error. In narrowband speech coding, the number of LSF parameters is 10, while it is 20 in the wideband MPEG-4 CELP speech encoding mode. The number of bits used for LSF quantization is 22 for the narrowband case and 46 bits for the wideband scenario, which involves 25 bits used for quantizing the first ten LSF coefficients and 21 bits for the 10 remaining LSFs [41].

The procedure of spectral envelope interpolation can also be employed for interpolating both the LARs and LSFs. The **Interpolation** flag, together with the **LPC_Present** flag unambiguously describe, how the LPC coefficients of the current frame are derived. The associated functionalities are summarised in Table 11.5. Specifically, if the **Interpolation** flag is set to one, this implies that the LPC coefficients of the current 20 ms frame are calculated by using the LPC coefficients of the previous and next frames. This would mean in general the decoding of the current frame must be delayed by one frame. In order to avoid the latency of one frame delay at the decoder, the LPC coefficients of the next frame are enclosed in the current frame [41]. In this case, the **LPC_Present** flag is set. Since the LPC coefficients of the next frame are already present in the current frame, the next frame will contain no LPC information. When the **Interpolation** flag is zero and the **LPC_Present** flag is zero, the LPC parameters of the current frame are those received in the previous frame. When the **Interpolation** flag is zero and the **LPC_Present** flag is one, then the current frame is a complete frame and the LPC parameters received in the current frame belong to the current frame. Note that in order to maintain good subjective speech quality, it is not allowed to have consecutive frames without the LPC information. This means the **Interpolation** flag may not have a value of 1 in two successive frames.

11.3.4 Multi Pulse and Regular Pulse Excitation

In MPEG-4 CELP coding, the excitation vectors can be encoded either using the Multi-Pulse Excitation (MPE) [477] or Regular-Pulse Excitation (RPE) [478] techniques. MPE is the default mode used for narrowband speech coding while RPE is the default mode for wideband speech coding, due to its simplicity in comparison to the MPE technique.

In Analysis-by-Synthesis (AbS) based speech codecs, the excitation signal is represented by a linear combination of the adaptive codevector and the fixed codevector scaled by their respective gains. Each component of the excitation signal is chosen by an analysis-by-synthesis search procedure in order to ensure that the perceptually weighted error between the input sig-

Bit Rate Range (kbit/s)	Frame Length (ms)	No. subframes per frame	No. pulses per subframe
3.85 - 4.65	40	4	3...5
4.90 - 5.50	30	3	5...7
5.70 - 7.30	20	2	6...12
7.70 - 10.70	20	4	4...12
11.00 - 12.20	10	2	8...12

Table 11.6: Excitation configurations for narrowband MPE.

Bit Rate Range (kbit/s)	Frame Length (ms)	No. subframes per frame	No. pulses per subframe
10.9 - 13.6	20	4	5...11
13.7 - 14.1	20	8	3...10
14.1 - 17.0	10	2	5...11
21.1 - 23.8	10	4	3...10

Table 11.7: Excitation configurations for wideband MPE

nal and the reconstructed signal is minimized [367]. The adaptive codebook parameters are constituted by the closed-loop delay and gain. The closed-loop delay is selected with the aid of a focussed search in the range around the estimated open-loop delay. The adaptive codevector is generated from a block of the past excitation signal samples associated with the selected closed-loop delay. The fixed codevector contains several non-zero excitation pulses. The excitation pulse positions obey an algebraic structure [367, 479]. In order to improve the achievable performance, after determining several sets of excitation pulse position candidates, a combined search based on the amalgamation of the excitation pulse position candidates and the pulse amplitudes is carried out.

For narrowband speech coding utilising Multi-Pulse Excitation (MPE) [477], the bit rate can vary from 3.85 to 12.2 kbit/s when using different configurations based on varying the frame length, the number of subframes per frame, and the number of pulses per subframe. These different configurations are shown in Table 11.6 and Table 11.7 for narrowband MPE and wideband MPE, respectively.

On the other hand, Regular Pulse Excitation (RPE) [478, 480] enables implementations having significantly lower encoder complexity and only slightly reduced compression efficiency. The RPE principle is used in wideband speech encoding, replacing MPE as the default mode and supporting bit rates between 13 and 24 kbit/s. RPE employs fixed pulse spacing, which implies that the distance of subsequent excitation pulses in the fixed codebook is fixed. This reduces the codebook search complexity required for obtaining the best indices during the analysis-by-synthesis procedure.

Having introduced the most important speech and audio coding modes of the MPEG-4 codec, let us now characterize its performance in the next section.

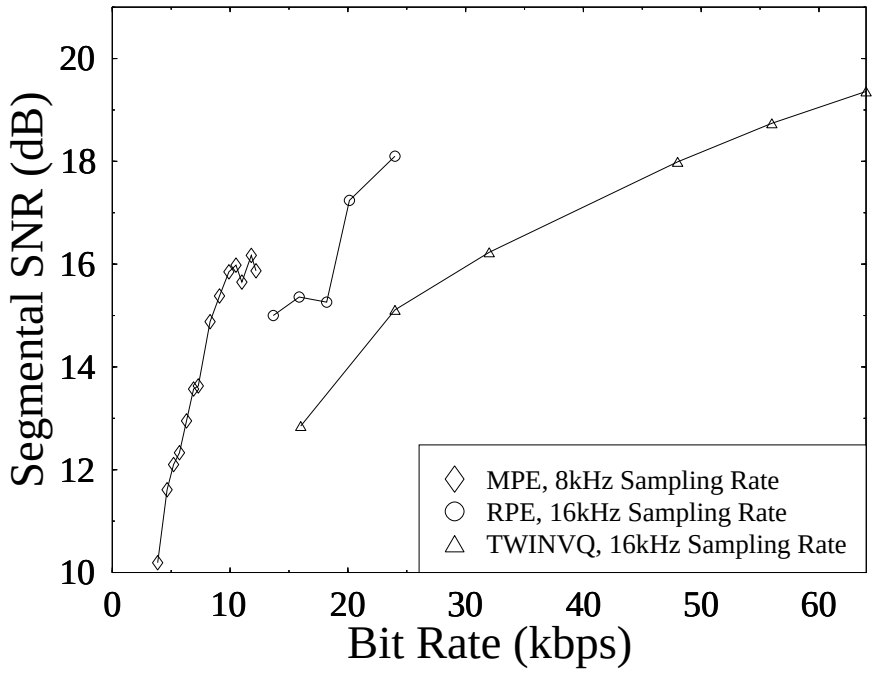

Figure 11.20: Segmental SNR performance for the encoding of speech signals, with respect to three different MPEG-4 coding modes, where the MPE codec and the RPE codec are employed at the sampling rates of 8 kHz and 16 kHz, respectively, while the TWINVQ audio codec of Section 11.2.9 operates at 16 kHz sampling rate.

11.4 MPEG-4 Codec Performance

Figure 11.20 shows the achievable Segmental SNR performance of the MPEG-4 codec at various bit rates applying various speech and audio coding modes. The MPE speech codec mode has been applied for bit rates between 3.85 kbit/s and 12.2 kbit/s for encoding narrow-band speech while the RPE codec in the CELP 'toolbox' is employed for wideband speech encoding spans from 13 kbit/s to 24 kbit/s. The TWINVQ audio codec of Section 11.2.9 was utilised for encoding music signals for bit rates of 16 kbit/s and beyond. In Figure 11.20, the codecs were characterized in terms of their performance, when encoding speech signals. As expected, the Segmental SNR increases upon increasing the bit rate. When the RPE codec mode is used, the wideband speech quality is improved in terms of both the objective Segmental SNR measure and the subjective quality. For the case of TWINVQ codec mode of Section 11.2.9, the Segmental SNR increases near-linearly with the bit rates. It is worth noting in Figure 11.20, that the RPE codec mode outperformed the TWINVQ codec mode over

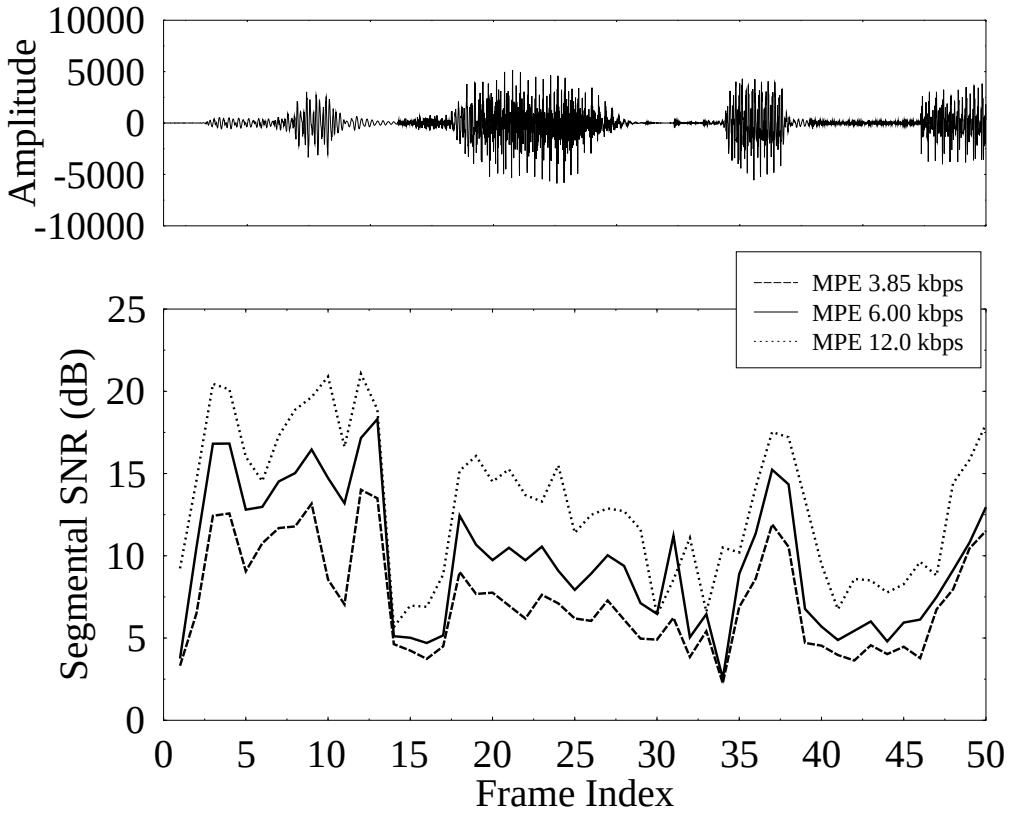

Figure 11.21: Segmental SNR performances versus frame index for three different bit rates of 3.85, 6.0 and 12.0 kbit/s, using the CELP tool in MPEG-4 Audio at a sampling rate of 8 kHz for the speech file of five.bin.

its entire bit rate range in the context of wideband speech encoding. This is because the RPE scheme is a dedicated speech codec while the TWINVQ codec is a more general audio codec, but capable of also encoding speech signals.

Figure 11.21 displays the achievable Segmental SNR performance versus frame index for the three different narrowband speech coding bit rates of 3.85, 6.0 and 12.0 kbit/s, using the MPE tool of the MPEG-4 Audio standard. The MPE tool offers the option of multi-rate coding, which is very useful in adaptive transmission schemes that can adapt the source bit rate according to the near-instantaneous channel conditions.

The performance of various codecs of the MPEG-4 toolbox is used for the encoding of music signals is shown in Figure 11.22 at a sampling rate of 16 kHz. We observe that as expected, the TWINVQ codec of Section 11.2.9 performed better, than the CELP codec when encoding music signals. The difference in Segmental SNR performance can be as high as 2

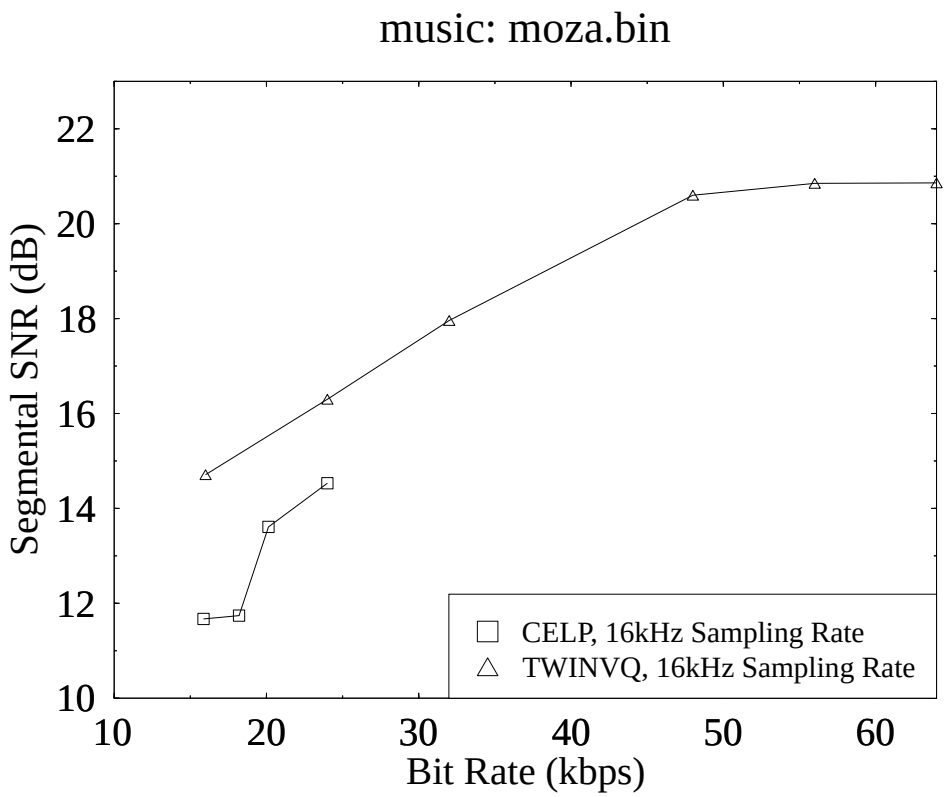

Figure 11.22: Comparing Segmental SNR performances for two different codecs - CELP and TWINVQ codecs at 16 kHz sampling rate, for coding of music file of moza.bin.

dB at the same bit rate.

11.5 MPEG-4 Space-Time Block Coded OFDM Audio Transceiver

[1] The Third Generation (3G) mobile communications standards [481] are expected to provide a wide range of bearer services, spanning from voice to high-rate data services, supporting rates of at least 144 kbit/s in vehicular, 384 kbit/s in outdoor-to-indoor and 2 Mbit/s in indoor as well as in picocellular applications.

In an effort to support such high rates, the bit/ symbol capacity of band-limited wireless channels can be increased by employing multiple antennas [482]. The concept of Space-Time

[1]This section is based on How, Liew and Hanzo: An MPEG-4 Space-Time OFDM Audio Transceiver, submitted to IEEE Proceedings of VTC, New Jersey, USA, 2001 and it was based on collaborative research with the co-authors.

Trellis Codes (STTCs) was proposed by Tarokh, Seshadri and Calderbank [483] in 1998. By jointly designing the FEC, modulation, transmit diversity and optional receive diversity scheme, they increased the effective bits/symbol (BPS) throughput of band-limited wireless channels, given a certain channel quality. A few months later, Alamouti [50] invented a low-complexity Space-Time Block Code (STBC), which imposes a significantly lower complexity at the cost of a slight performance degradation. Alamouti's invention motivated Tarokh *et al.* [484, 485] to generalise Alamouti's scheme to an arbitrary number of transmitter antennas. Then, Tarokh *et al.*, Bauch *et al.* [486], Agrawal [487], Li *et al.* [488] and Naguib *et al.* [489] extended the research of space-time codes from considering narrow-band channels to dispersive channels [490]. The benefits of space time coding in terms of mitigating the effects of channel fading are substantial and hence they were optionally adopted in the forthcoming 3G cellular standards [491].

In recent years substantial advances have been made in the field of Orthogonal Frequency Division Multiplexing (OFDM), which was first proposed by Chang in his 1966 paper [492]. Research in OFDM was revived amongst others by Cimini in his often cited paper [493] and the field was further advanced during the nineties, with a host of contributions documented for example in [494]. In Europe, OFDM has been favoured for both Digital Audio Broadcasting (DAB) and Digital Video Broadcasting (DVB) [495, 496] as well as for high-rate Wireless Asynchronous Transfer Mode (WATM) systems due to its ability to combat the effects of highly dispersive channels [497]. Most recently OFDM has been also proposed for the downlink of high-rate wireless Internet access [498].

At the time of writing we are witnessing the rapid emergence of intelligent multi-mode High-Speed Downlink Packet Access (HSDPA) style mobile speech and audio communicators [339, 371, 499], that can adapt their parameters in response to rapidly changing propagation environments. Simultaneously, significant efforts have been dedicated to researching multi-rate source coding, which are required by the near-instantaneously adaptive transceivers [500]. The recent GSM Adaptive Multi-Rate (AMR) standardization activities have prompted significant research interests in invoking the AMR mechanism in half-rate and full-rate channels [28]. Recently ETSI also standardized the wideband AMR (AMR-WB) speech codec [337] for the GSM system, which provides a high speech quality due to representing the extra audio bandwidth of 7 kHz, instead of the conventional 3.1 kHz bandwidth. Finally, the further enhanced AMR-WB+ audio- and speech codec was detailed in Section 9.7.

The standardization activities within the framework of the MPEG-4 audio coding initiative [501] have also reached fruition, supporting the transmission of natural audio signals, including the representation of synthetic audio, such as Musical Instrument Digital Interface (MIDI) [48] and Text-to-Speech (TTS) systems [42]. A wide ranging set of bit rates spanning from 2 kbit/s per channel up to 64 kbit/s per channel are supported by the MPEG-4 audio codec.

Against this backcloth, in this section the underlying trade-offs of using the multi-rate MPEG-4 TWINVQ audio encoder of Section 11.2.9, in conjunction with a turbo-coded [502] and space-time coded [483], reconfigurable BPSK/QPSK/16QAM OFDM system [51] are investigated, in order to provide an attractive system design example.

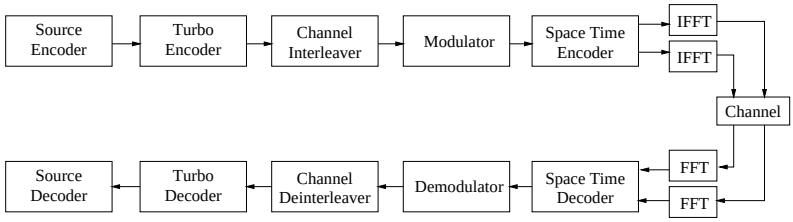

Figure 11.23: Schematic overview of the turbo-coded and space-time coded OFDM system.

11.5.1 System Overview

Figure 11.23 shows the schematic of the turbo-coded and space-time coded OFDM system. The source bits generated by the MPEG-4 TWINVQ encoder [41] are passed to the turbo encoder using the half-rate, constraint length three turbo convolutional encoder TC(2,1,3), employing an octal generator polynomial of (7,5). The encoded bits were channel interleaved and passed to the modulator. The choice of the modulation scheme to be used by the transmitter for its next OFDM symbol is determined by the channel quality estimate of the receiver based on the current OFDM symbol. Here, perfect channel quality estimation and perfect signalling of the required modem modes were assumed. In order to simplify the task of signalling the required modulation modes from receiver A to transmitter B, we employed the subband-adaptive OFDM transmission scheme proposed by Keller *et al.* [51]. More specifically, the total OFDM symbol bandwidth was divided into equi-width subbands having a similar channel quality, where the same modem mode was assigned. The modulated signals were then passed to the encoder of the space-time block code G_2 [50], which employs two transmitters and one receiver. The space-time encoded signals were OFDM modulated and transmitted by the corresponding antennas.

The received signals were OFDM demodulated and passed to the space-time decoders. Logarithmic Maximum Aposteriori (Log-MAP) decoding [503] of the received space-time signals was performed, in order to provide soft-outputs for the TC(2,1,3) turbo decoder. The received bits were then channel deinterleaved and passed to the TC decoder, which again, employs the Log-MAP decoding algorithm. The decoded bits were finally passed to the MPEG-4 TWINVQ decoder for obtaining the reconstructed audio signal.

11.5.2 System parameters

Table 11.8 and 11.9 gives an overview of the proposed system's parameters. The transmission parameters have been partially harmonised with those of the TDD-mode of the Pan-European UMTS system [491]. The sampling rate is assumed to be 1.9 MHz, leading to a 1024 subcarrier OFDM symbol. The channel model used was the four-path COST 207 Typical Urban (TU) Channel Impulse Response (CIR) [409], where each impulse was subjected to independent Rayleigh fading having a normalised Doppler frequency of $2.25 \cdot 10^{-6}$, corresponding to a pedestrian scenario at a walking speed of 3mph. The channel impulse response is shown in Figure 11.24.

The channel encoder is a convolutional constituent coding based turbo encoder [502], em-

System Parameters	Value
Carrier Frequency	1.9 GHz
Sampling Rate	3.78 MHz
Channel	
Impulse Response	COST207
Normalised Doppler Frequency	$2.25 \cdot 10^{-6}$
OFDM	
Number of Subcarriers	1024
OFDM Symbols/Packet	1
OFDM Symbol Duration	(1024+64) x $1/(3.78 \cdot 10^6)$
Guard Period	64 samples
Modulation Scheme	Fixed Modulations
Space Time Coding	
Number of transmitters	2
Number of receivers	1
Channel Coding	Turbo Convolutional
Constraint Length	3
Code Rate	0.5
Generator Polynomials	7, 5
Turbo Interleaver Length	464/928/1856/2784
Decoding Algorithm	Log MAP
Number of Iterations	8
Source Coding	MPEG-4 TWINVQ
Bit Rates (kbit/s)	16 - 64
Audio Frame Length (ms)	23.22
Sampling Rate (kHz)	44.1

Table 11.8: System Parameters

ploying block turbo interleavers and a pseudo-random channel interleaver. Again, the constituent Recursive Systematic Convolutional (RSC) encoder employs a constraint length of 3 and the octal generator polynomial of (7,5). Eight iterations are performed at the decoder, utilising the MAP-algorithm and the Log-Likelihood Ratio (LLR) soft inputs provided by the demodulator.

The MPEG-4 TWINVQ audio coder has been chosen for this system, which can be programmed to operate at bit rates between 16 and 64 kbit/s. It provides a high audio quality at an adjustable bit rate and will be described in more depth in the next section.

11.5.3 Frame Dropping Procedure

For completeness, we investigated the bit sensitivity of the TWINVQ codec. A high robustness against bit errors, inflicted by wireless channels is an important criterion for the design of a communication system. A commonly used approach in quantifying the sensitivity of a

Data + Parity Bits	928	1856	3712
Source Coded Bits/Packet	372	743	1486
Source Coding Bit Rate (kbit/s)	16	32	64
Modulation Mode	BPSK	QPSK	16QAM
Minimum Channel SNR for 1% FER (dB)	4.3	7.2	12.4
Minimum Channel SNR for 5% FER (dB)	2.7	5.8	10.6

Table 11.9: System parameters

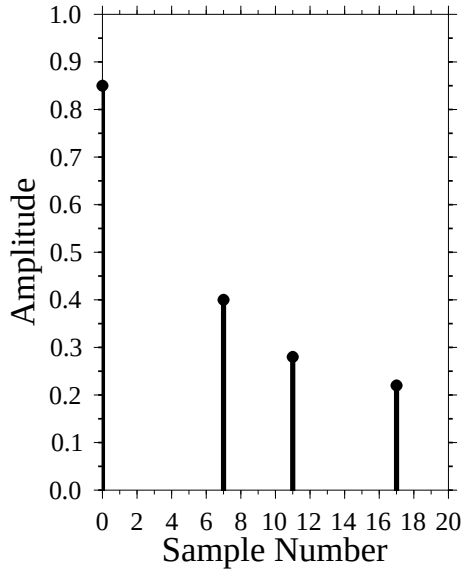

Figure 11.24: COST207 channel impulse response [409].

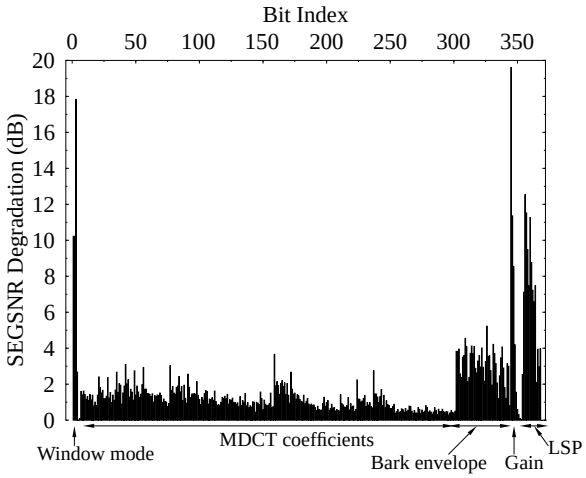

Figure 11.25: SEGSNR degradation against bit index using MPEG-4 TWINVQ at 16 kbit/s. The corresponding bit allocation scheme was given in Table 11.1.

given bit is to invert this bit consistently in every audio frame and to evaluate the associated Segmental SNR (SEGSNR) degradation [371]. Figure 11.25 shows the bit error sensitivity of the MPEG-4 TWINVQ encoder of Section 11.2.9 at 16 kbit/s. This figure shows that the bits representing the gain factors (bit 345-353), the LSF parameters (bit 354-372), and the Bark-envelope (bit 302-343) are more sensitive to channel errors, compared to the bits representing the MDCT coefficients (bit 7-301). The bits signalling the window mode used are also very sensitive to transmission errors and hence have to be well protected. The window modes were defined in Section 11.2.

In Section 7.13.5.2 we studied the benefits of invoking multi-class embedded error correction coding assigned to the narrowband AMR speech codec, while in Section 10 in the context of the AMR-WB codec. By contrast, in the wideband MPEG-4 TWINVQ system studied here erroneously received audio frames are dropped and replaced by the previous audio frame, since the system is aiming for maintaining a high audio quality and the error-infested audio frames would result in catastrophic inter-frame error propagation. Hence the system's audio quality is determined by the tolerable transmission Frame Error Rate (FER), rather than by the BER. In order to determine the highest FER that can be tolerated by the MPEG-4 TWINVQ codec, it was exposed to random frame dropping and the associated SEGSNR degradation as well as the informally assessed perceptual audio degradation was evaluated. The corresponding SEGSNR degradation is plotted in Figure 11.26. Observe in the figure that at a given FER the higher rate modes suffer from a higher SEGSNR degradation. This is because their audio SEGSNR is inherently higher and hence for example obliterating one frame in 100 frames inevitably reduces the average SEGSNR more dramatically. We found that the associated audio quality expressed in terms of Segmental SNR (SEGSNR) degradation was deemed to be perceptually objectionable for frame error rates in excess of 1%. Again, frame dropping was preferred, which was found to be more beneficial

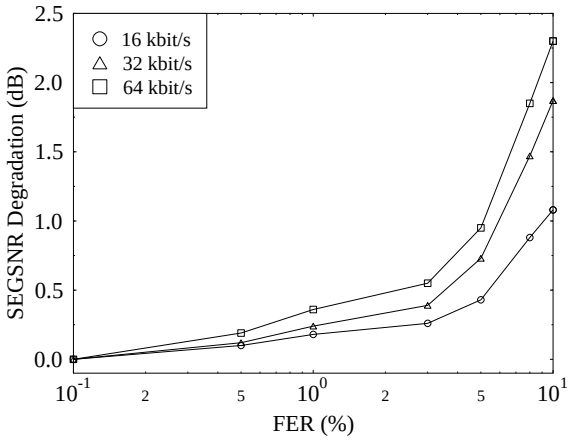

Figure 11.26: SEGSNR degradation against FER for the MPEG-4 TWINVQ codec of Section 11.2.9, at bit rates of 16, 32 and 64 kbit/s. The SEGSNR degradation values were obtained, in conjunction with the employment of frame dropping.

in audio quality terms, than retaining corrupted audio frames.

For the sake of completeness, Figure 11.27 shows the SEGSNR degradation when inflicting random bit errors but retaining the corrupted audio frames. As expected, the highest bit rate mode of 64 kbit/s suffered the highest SEGSNR degradation upon increasing the BER, since a higher number of bits per frame was corrupted by errors, which degraded the audio quality more considerably.

11.5.4 Space-Time Coding

Traditionally, the most effective technique of combating fading has been the exploitation of diversity [483]. Diversity techniques can be divided into three broad categories, namely temporal diversity, frequency diversity and spatial diversity. Temporal and frequency diversity schemes [50] introduce redundancy in the time and/or frequency domain, which results in a loss of bandwidth efficiency. Examples of spatial diversity are constituted by multiple transmit- and/or receive-antenna based systems [483]. Transmit-antenna diversity relies on employing multiple antennas at the transmitter and hence it is more suitable for downlink transmissions, since having multiple transmit antennas at the base station is certainly feasible. By contrast, receive-antenna diversity employs multiple antennas at the receiver for acquiring multiple copies of the transmitted signals, which are then combined in order to mitigate the channel-induced fading.

Space time coding [50,483] is a specific form of transmit-antenna diversity, which aims for usefully exploiting the multipath phenomenon experienced by signals propagating through the dispersive mobile channel. This is achieved by combining multiple transmission antennas in conjunction with appropriate signal processing at the receiver, in order to provide diversity

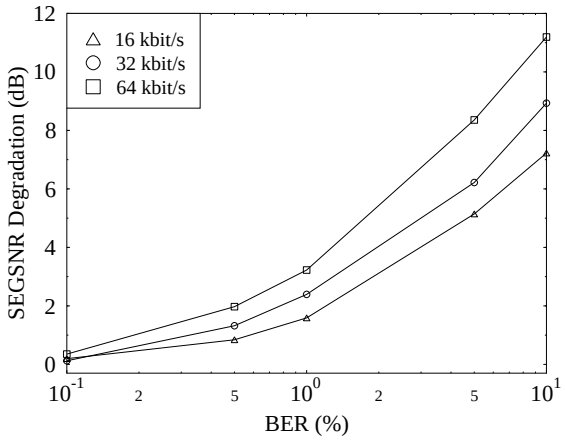

Figure 11.27: SEGSNR degradation against BER for the MPEG-4 TWINVQ codec of Section 11.2.9, at bit rates of 16, 32 and 64 kbit/s.

and coding gain in comparison to uncoded single-antenna scenarios [489].

In the system investigated, we employ a two-transmitter and one-receiver configuration, in conjunction with turbo channel coding [502]. In Figure 11.28, we show the instantaneous channel SNR experienced by the 512-subcarrier OFDM modem for a one-transmitter, one-receiver scheme and for the space time block code G_2 [50] using two transmitters and one receiver for transmission over the COST207 channel. The average channel SNR was 10 dB. We can see in Figure 11.28 that the variation of the instantaneous channel SNR for a one-transmitter, one-receiver scheme is severe. The instantaneous channel SNR may become as low as 4 dB due to the deep fades inflicted by the channel. On the other hand, we can see that for the space-time block code G_2 using one receiver the variation of the instantaneous channel SNR is less severe. Explicitly, by employing multiple transmit antennas in Figure 11.28, we have significantly reduced the depth of the channel fades. Whilst space-time coding endeavours to mitigate the fading-related time- and frequency-domain channel-quality fluctuations at the cost of increasing the transmitter's complexity, adaptive modulation attempts to accommodate these channel quality fluctuations, as it will be outlined in the next section.

11.5.5 Adaptive Modulation

In order to accommodate the time- and frequency-domain channel quality variations seen in case of the 1Tx 1Rx scenario of Figure 11.28, the employment of a multi-mode system is desirable, which allows us to switch between a set of different source- and channel encoders as well as various transmission parameters, depending on the instantaneous channel quality [51].

In the proposed system, we have defined three operating modes, which correspond to the uncoded audio bit rates of 16, 32 and 64 kbit/s. This corresponds to 372, 743 and 1486 bits

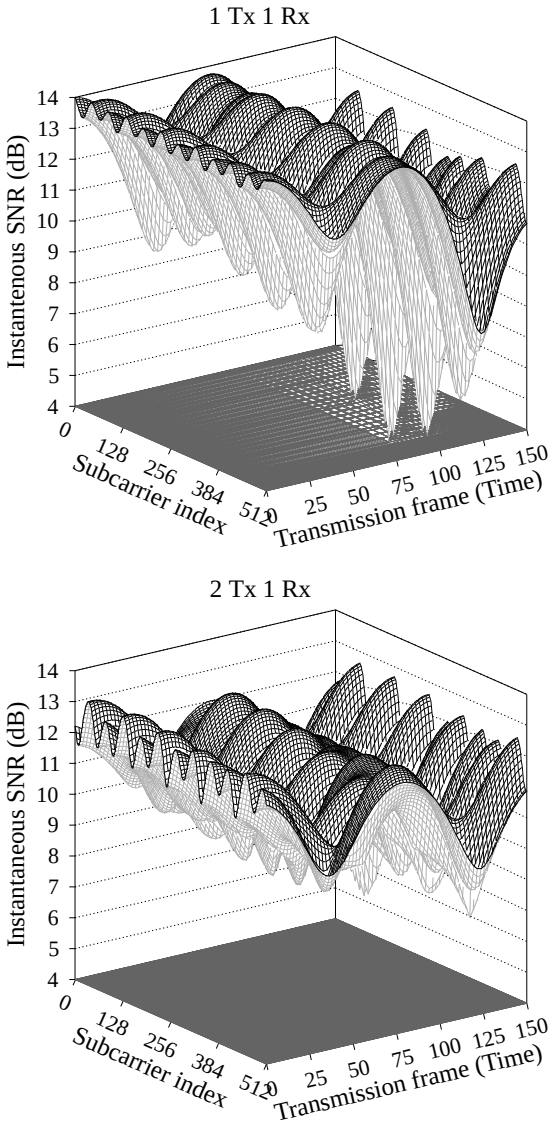

Figure 11.28: Instantaneous channel SNR of 512-subcarrier OFDM symbols for one-transmitter one-receiver (1Tx 1Rx) and for the space-time block code using two-transmitter one-receiver (2Tx 1Rx).

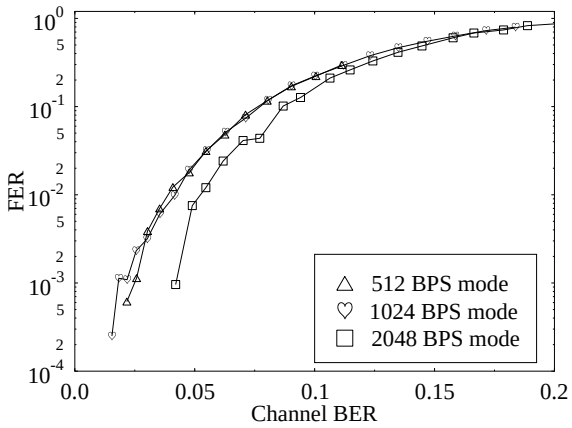

Figure 11.29: FER against channel BER performance of the adaptive OFDM modem conveying 512, 1024 and 2048 BPS for transmission over the channel model of Figure 11.24.

per 23.22 ms audio frame. In conjunction with half-rate channel coding and also allowing for check sums and signalling overheads, the number of transmitted turbo coded bits per OFDM symbol is 928, 1856 and 3712 for the three source-coded modes, respectively. Again, these bit rates are also summarised in Table 11.9. Each transmission mode uses a different modulation scheme, depending on the instantaneous channel conditions. It is beneficial, if the transceiver can drop its source rate, for example from 64 kbit/s to 32 kbit/s and invoke QPSK modulation instead of 16QAM, while maintaining the same bandwidth. Hence, during good channel conditions the higher throughput, higher audio quality but less robust modes of operation can be invoked, while the more robust but lower audio quality BPSK/16kbit/s mode can be applied during degrading channel conditions.

Figure 11.29 shows the FER observed for all three modes of operation, namely for the 512, 1024 and 2048 versus the channel BER that was predicted by the OFDM receiver during the channel quality estimation process. Again, the rationale behind using the FER, rather than the BER for estimating the expected channel quality of the next transmitted OFDM symbol is, because the MPEG-4 audio codec has to drop the turbo-decoded received OFDM symbols, which contained transmission errors. This is because corrupted audio packets would result in detrimental MPEG-4 decoding error propagation and audio artifacts. A FER of 1% was observed for an estimated input bit error rate of about 4% for the 16 and 32 kbit/s modes, while a BER of over 5% was tolerable for the 64 kbit/s mode. This was, because the number of bits per OFDM symbol was quadrupled in the 16QAM mode over which turbo interleaving was invoked compared to the BPSK mode. The quadrupled interleaving length substantially increased the turbo codec's performance.

In Figure 11.30, we show our Bits Per Symbol (BPS) throughput performance comparison between the subband-adaptive and fixed mode OFDM modulation schemes. From the figure we can see that at a low BPS throughput the adaptive OFDM modulation scheme outper-

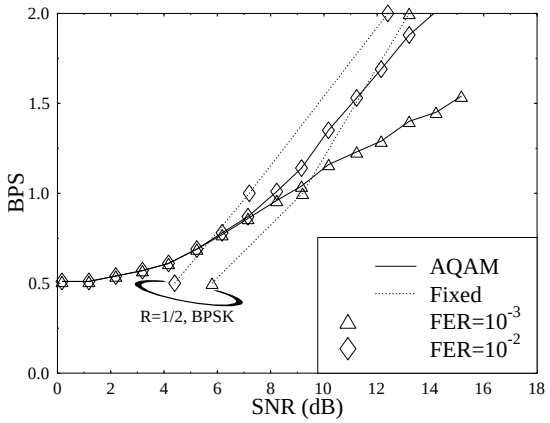

Figure 11.30: BPS performance comparison between the adaptive and fixed-mode OFDM modulation schemes, when using space-time coding, for transmission over the channel model of Figure 11.24.

forms the fixed OFDM modulation scheme. However, as the BPS throughput of the system increases, the fixed modulation schemes become preferable. This is, because adaptive modulation is advantageous, when there are high channel quality variations in the one-transmitter, one receiver scheme. However, we have shown in Figure 11.28 that the channel quality variations have been significantly reduced by employing two $\mathbf{G_2}$ space-time transmitters. Therefore, the advantages of adaptive modulation eroded due to the reduced channel quality variations in the space-time coded system. As a consequence, two different-complexity system design principles can be proposed. The first system is the lower-complexity one-transmitter, one receiver scheme, which mitigates the severe variation of the channel quality by employing subband adaptive OFDM modulation. By contrast, we can design a more complex $\mathbf{G_2}$ space-time coded system, which employs fixed modulation schemes, since no substantial benefits accrue from employing adaptive modulation, once the fading-induced channel-quality fluctuations have been sufficiently mitigated by the $\mathbf{G_2}$ space-time code. In the remainder of this section, we have opted for investigating the performance of the more powerful space-time coded system, requiring an increased complexity.

11.5.6 System Performance

As mentioned before, the detailed subsystem parameters used in our space-time coded OFDM system are listed in Table 11.8. Again, the channel impulse response profile used was the COST 207 Typical Urban (TU) channel [409] having four paths and a maximum dispersion of 4.5 μs, where each path was faded independently at a Doppler frequency of $2.25 \cdot 10^{-6}$ Hz.

The BER is plotted versus the channel SNR in Figure 11.31 for the three different fixed

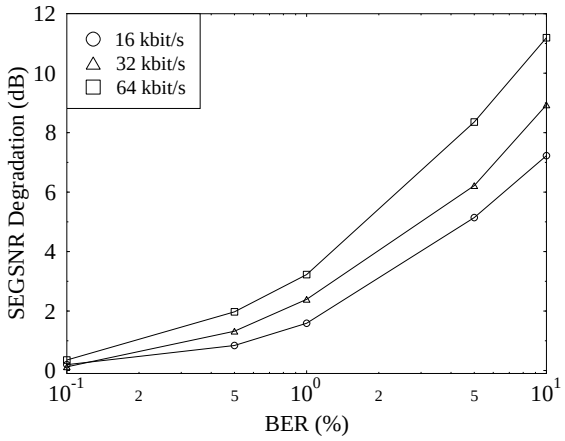

Figure 11.31: BER against Channel SNR performance of the fixed-mode OFDM transceiver of Table 11.8 in conjunction with and without space time coding, in comparison to the conventional one-transmitter, one-receiver benchmarker for transmission over the channel model of Figure 11.24.

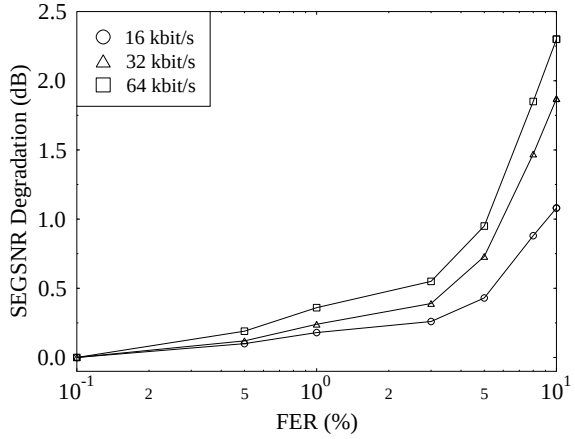

Figure 11.32: FER against Channel SNR performance of the fixed-mode OFDM transceiver of Table 11.8 in conjunction with and without space time coding, in comparison with the conventional one-transmitter, one-receiver benchmarker for transmission over the channel model of Figure 11.24.

modes of operation conveying 512, 1024 or 2048 bits per OFDM symbol both with and without space-time coding. The employment of space time coding improved the system's performance significantly, giving an approximately 3 dB channel SNR improvement at a BER of 1%. As expected, the lowest throughput BPSK/ 16kbit/s mode was more robust in BER terms, than the QPSK/32kbit/s and the 16QAM /64kbit/s configurations, albeit delivering a lower audio quality. Similar results were obtained in terms of FER versus the channel SNR, which are displayed in Figure 11.32, indicating that the most robust BPSK/16kbit/s scheme performed better than the QPSK/32kbit/s and 16QAM/64kbit/s configurations, albeit at a lower audio quality.

The overall SEGSNR versus channel SNR performance of the proposed audio transceiver is displayed in Figure 11.33, again, employing $\mathbf{G_2}$ space-time coding using two transmitters and one receiver. The lower-complexity benchmarker using the conventional one-transmitter, one-receiver scheme was also characterized in the figure. We observe again that the employment of space time coding provides a substantial improvement in terms maintaining an error free audio performance. Specifically, an SNR advantage of 4 dB was recorded compared to the conventional lower-complexity one-transmitter, one-receiver benchmarker for all three modulation modes. Furthermore, focussing on the three different operating modes using space-time coding, namely on the curves drawn in continuous lines, the 16QAM/64kbit/s mode was shown to outperform the QPSK/32kbit/s scheme in terms of both objective and subjective audio quality for channel SNRs in excess of about 10 dB. At a channel SNR of about 9 dB, where the 16QAM and QPSK SEGSNR curves cross each other in Figure 11.33, it is preferable to invoke the inherently lower audio quality, but unimpaired QPSK mode of operation. Similarly, at a Channel SNR around 5 dB, when the QPSK/32kbit/s scheme's performance starts to degrade, it is better to invoke the unimpaired BPSK/ 16kbit/s mode of operation, in order to avoid the channel-induced audio artifacts.

11.6 Turbo-Detected Space-Time Trellis Coded MPEG-4 Audio Transceivers

N. S. Othman, S. X. Ng and L. Hanzo

11.6.1 Motivation and Background

In this section a jointly optimised turbo transceiver capable of providing unequal error protection is proposed for employment in an MPEG-4 coded audio transceiver. The transceiver advocated consists of Space-Time Trellis Coding (STTC), Trellis Coded Modulation (TCM) and two different-rate Non-Systematic Convolutional codes (NSCs) used for unequal error protection. A benchmarker scheme combining STTC and a single-class protection NSC is used for comparison with the proposed scheme. The audio performance of the both schemes will be evaluated when communicating over uncorrelated Rayleigh fading channels. We will demosntrate that the proposed unequal protection turbo-transceiver scheme requires about two dBs lower transmit power than the single-class turbo benchmarker scheme in the context of the MPEG-4 audio transceiver, when aiming for an effective throughput of 2 bits/symbol, while exhibiting a similar decoding complexity.

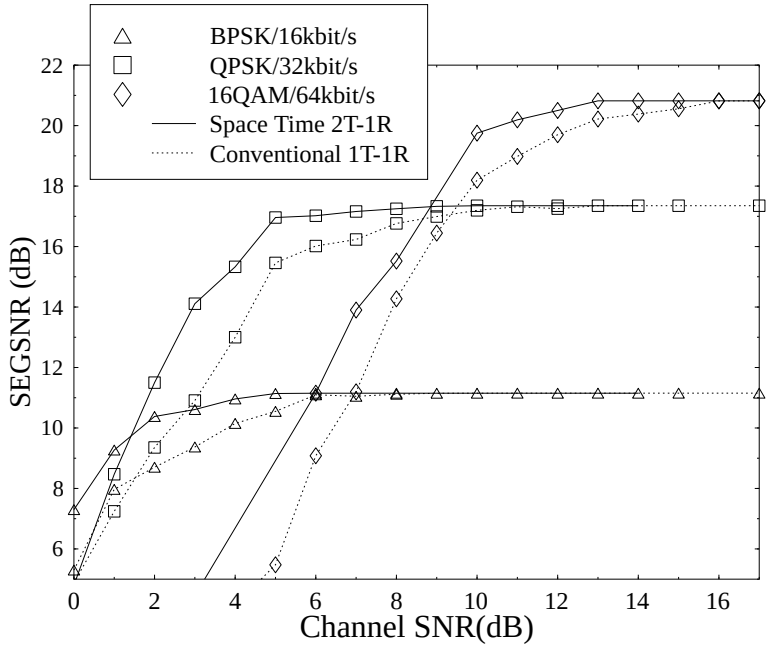

Figure 11.33: SEGSNR against channel SNR of the MPEG-4 TWINVQ based fixed mode OFDM transceiver in conjunction with and without space time coding, in comparison to the conventional one-transmitter, one-receiver benchmarker.

The previously characterized MPEG-4 standard [504, 505] defines a comprehensive multimedia content representation scheme that is capable of supporting numerous applications - such as streaming multimedia signals over the internet/intranet, content-based storage and retrieval, digital multimedia broadcast or mobile communications. The audio-related section of the MPEG-4 standard [506] defines audio codecs covering a wide variety applications - ranging from narrowband low-rate speech to high quality multichannel audio, and from natural sound to synthesized sound effects as a benefit of its object-based approach used for representing the audio signals.

The MPEG-4 General Audio (GA) encoder is capable of compressing arbitrary natural audio signals. One of the key components of the MPEG-4 GA encoder is the Time/Frequency (T/F) compression scheme constituted by the Advanced Audio Coding (AAC) and Transform based Weighted Vector Quantization (TwinVQ), which is capable of operating at bitrates ranging from 6 kbit/s to broadcast quality audio at 64 kbit/s [504].

The MPEG-4 T/F codec is based on the MPEG-2 AAC standard, extended by a number of additional functionalities, such as Perceptual Noise Substitution (PNS) and Long Term Prediction (LTP) for enhancing the achievable compression performance, and combined with the TwinVQ for operation at extremely low bit rates. Another important feature of this codec is its

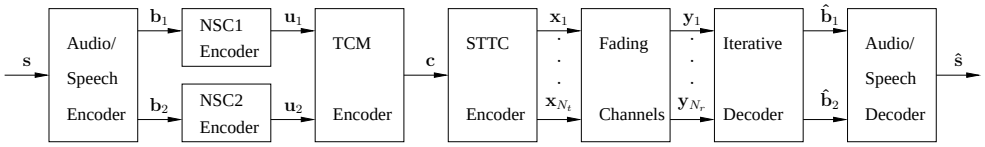

Figure 11.34: Block diagram of the serially concatenated STTC-TCM-2NSC assisted MPEG-4 audio scheme. The notations $\mathbf{s}$, $\hat{\mathbf{s}}$, $\mathbf{b}_i$, $\hat{\mathbf{b}}_\mathbf{i}$, $\mathbf{u}_i$, $\mathbf{c}$, $\mathbf{x}_j$ and $\mathbf{y}_k$ denote the vector of the audio source symbol, the estimate of the audio source symbol, the class-i audio bits, the estimates of the class-i audio bits, the encoded bits of class-i NSC encoders, the TCM coded symbols, the STTC coded symbols for transmitter j and the received symbols at receiver k, respectively. Furthermore, N_t and N_r denote the number of transmitters and receivers, respectively. The symbol-based channel interleaver between the STTC and TCM schemes as well as the two bit-based interleavers at the output of NSC encoders are not shown for simplicity. The iterative decoder seen at the right is detailed in Figure 11.35.

robustness against transmission errors in error-prone propagation channels [507]. The error resilience of the MPEG-4 T/F codec is mainly attributed to the so-called Virtual Codebook tool (VCB11), Reversible Variable Length Coding tool (RVLC) and Huffman Codeword Reordering tool (HCR) [507, 508], which facilitate the integration of the MPEG-4 T/F codec into wireless systems.

In this study the MPEG-4 audio codec was incorporated in a sophisticated unequal-protection turbo transceiver using joint coding and modulation as inner coding, twin-class convolutional outer coding as well as space time coding based spatial diversity. Specifically, maximal minimum distance Non-Systematic Convolutional codes (NSCs) [509, p. 331] having two different code-rates were used as outer encoders for providing unequal audio protection. On one hand, Trellis Coded Modulation (TCM) [510–512] constitutes a bandwidth-efficient joint channel coding and modulation scheme, which was originally designed for transmission over Additive White Gaussian Noise (AWGN) channels. On the other hand, Space-Time Trellis Coding (STTC) [511, 513] employing multiple transmit and receive antennas is capable of providing spatial diversity gain. When the spatial diversity order is sufficiently high, the channel's Rayleigh fading envelope is transformed to a Gaussian-like near-constant envelope. Hence, the benefits of a TCM scheme designed for AWGN channels will be efficiently exploited, when TCM is concatenated with STTC.

We will demonstrate that significant iteration gains are attained with the aid of the proposed turbo transceiver. The section is structured as follows. In Section 11.6.2 we describe the MPEG-4 audio codec, while in Section 11.6.3 the architecture of the turbo transceiver is described. We elaborate further by characterising the achievable system performance in Section 11.6.4 and conclude in Section 11.6.5.

11.6.2 Audio Turbo Transceiver Overview

As mentioned above, the MPEG-4 AAC is based on time/frequency audio coding, which provides redundancy reduction by exploiting the correlation between subsequent audio samples of the input signal. Furthermore, the codec uses perceptual modelling of the human auditory

system for masking the quantisation distortion of the encoded audio signals by allowing more distortion in those frequency bands, where the signal exhibits higher energy peaks and vice versa [507, 508].

The MPEG-4 AAC is capable of providing an attractive audio quality versus bitrate performance, yielding high-fidelity audio reconstruction for bit rates in excess of 32 kbit/s per channel. In the proposed wireless system the MPEG-4 AAC is used for encoding the stereo audio file at a bit rate of 48 kbit/s. The audio input signal was sampled at 44.1 kHz and hence results in an audio framelength of 23.22 ms, which corresponds to 1024 audio input samples. The compressed audio information is formatted into a packetized bitstream, which conveyed one audio frame. In our system, the average transmission frame size is approximately 1116 bits per frame. The audio Segmental Signal to Noise Ratio (SegSNR) of this configuration was found to be S_0 = 16.28dB, which gives a transparent audio quality.

It is well recognised that in highly compressed audio bitstreams a low bit error ratio (BER) may lead to perceptually unacceptable distortion. In order to prevent the complete loss of transmitted audio frames owing to catastrophic error propagation, the most sensitive bits have to be well protected from channel errors. Hence, in the advocated system Unequal Error Protection (UEP) is employed, where the compressed audio bitstream was partitioned into two sensitivity classes. More explicitly, an audio bit, which resulted in a SegSNR degradation above 16 dB upon its corruption was classified into protection class-1. A range of different audio files were used in our work and the results provided are related to a 60 seconds long excerpt of Mozart's "Clarinet Concerto (2nd movement - Adagio)". From the bit sensitivity studies using this audio file as the source, we found that approximately 50% of the total number of MPEG-4 encoded bits falls into class-1.

At the receiver, the output of the turbo transceiver is decoded using the MPEG-4 AAC decoder. During the decoding process, the erroneously received audio frames were dropped and replaced by the previous error-free audio frame for the sake of avoiding an even more dramatic error-infested audio-quality degradation [514, 515].

11.6.3 The Turbo Transceiver

The block diagram of the serially concatenated STTC-TCM-2NSC turbo scheme using a STTC, a TCM and two different-rate NSCs as its constituent codes is depicted in Figure 11.40. Since the number of class-1 audio bits is approximately the same as that of the class-2 audio bits and there are approximately 1116 bits per audio frame, we protect the 558-bit class-1 audio sequence using a rate-R_1 NSC encoder and the 558-bit class-2 sequence using a rate-R_2 NSC encoder. Let us denote the turbo scheme as STTC-TCM-2NSC-1 when the NSC coding rates of $R_1 = k_1/n_1 = 1/2$ and $R_2 = k_2/n_2 = 3/4$ are used. Furthermore, when the NSC coding rates of $R_1 = 2/3$ and $R_2 = 3/4$ are used, we denote the turbo scheme as STTC-TCM-2NSC-2. The code memory of the class-1 and class-2 NSC encoders is $L_1 = 3$ and $L_2 = 3$, respectively. The class-1 and class-2 NSC coded bit sequences are interleaved by two separate bit interleavers, before they are fed to the rate-$R_3 = 3/4$ TCM [510–512] scheme having a code memory of $L_3 = 3$. Code termination was employed for the NSCs, TCM [510–512] and STTC codecs [511, 513]. The TCM symbol sequence is then symbol-interleaved and fed to the STTC encoder. We invoke a 16-state STTC scheme having a code memory of $L_4 = 4$ and $N_t = 2$ transmit antennas, employing $M = 16$-level Quadrature Amplitude Modulation (16QAM) [512]. The STTC employing

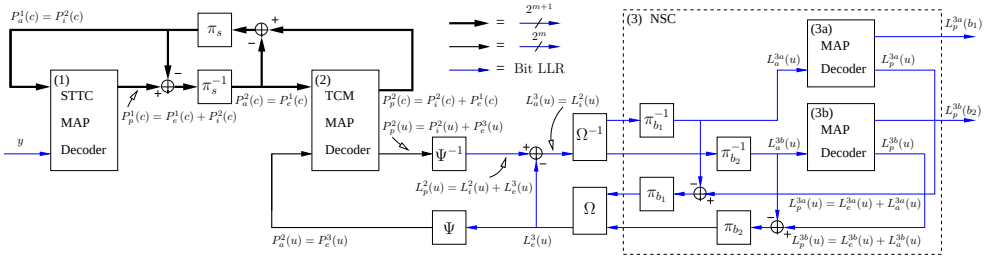

Figure 11.35: Block diagram of the STTC-TCM-2NSC turbo detection scheme seen at the right of Figure 11.40. The notations $\pi_{(s,b_i)}$ and $\pi^{-1}_{(s,b_i)}$ denote the interleaver and deinterleaver, while the subscript s denotes the symbol-based interleaver of TCM and the subscript b_i denotes the bit-based interleaver for class-i NSC. Furthermore, Ψ and Ψ^{-1} denote LLR-to-symbol probability and symbol probability-to-LLR conversion, while Ω and Ω^{-1} denote the parallel-to-serial and serial-to-parallel converter, respectively. The notation m denotes the number of information bits per TCM coded symbol. The thickness of the connecting lines indicates the number of non-binary symbol probabilities spanning from a single LLR per bit to 2^m and 2^{m+1} probabilities [516] ©IEE, 2004, Ng, Chung and Hanzo.

$N_t = 2$ requires one 16QAM-based termination symbol. The overall coding rate is given by $R_{s1} = 1116/2520 \approx 0.4429$ and $R_{s2} = 1116/2152 \approx 0.5186$ for the STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes, respectively. The effective throughput of the STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes is $\log_2(M)R_{s1} \approx 1.77$ Bits Per Symbol (BPS) and $\log_2(M)R_{s2} \approx 2.07$ BPS, respectively.

At the receiver, we employ $N_r = 2$ receive antennas and the received signals are fed to the iterative decoders for the sake of estimating the audio bit sequences in both class-1 and class-2, as seen in Figure 11.40. The STTC-TCM-2NSC scheme's turbo decoder structure is illustrated in Figure 11.35, where there are four constituent decoders, each labelled with a round-bracketed index. The Maximum A-Posteriori (MAP) algorithm [511] operating in the logarithmic-domain are employed by the STTC, TCM and the two NSC decoders, respectively. The notations $P(.)$ and $L(.)$ in Figure 11.35 denote the logarithmic-domain symbol probabilities and the Logarithmic-Likelihood Ratio (LLR) of the bit probabilities, respectively. The notations c, u and b_i in the round brackets $(.)$ in Figure 11.35 denote TCM coded symbols, TCM information symbols and the class-i audio bits, respectively. The specific nature of the probabilities and LLRs is represented by the subscripts a, p, e and i, which denote $a\ priori$, $a\ posteriori$, $extrinsic$ and $intrinsic$ information, respectively. The probabilities and LLRs associated with one of the four constituent decoders having a label of $\{1, 2, 3a, 3b\}$ are differentiated by the identical superscripts of $\{1, 2, 3a, 3b\}$. Note that the superscript 3 is used for representing the two NSC decoders of $3a$ and $3b$. The iterative turbo-detection scheme shown in Figure 11.35 enables an efficient information exchange between STTC, TCM and NSCs constituent codes for the sake of achieving spatial diversity gain, coding gain, unequal error proctection and a near-channel-capacity performance. The information exchange mechanism between each constituent decoders is detailed in [516].

For the sake of benchmarking the scheme advocated, we created a powerful benchmark

scheme by replacing the TCM and NSC encoders of Figure 11.40 by a single NSC codec having a coding rate of $R_0 = k_0/n_0 = 1/2$ and a code memory of $L_0 = 6$. We will refer to this benchmarker scheme as the STTC-NSC arrangement. All audio bits are equally protected in the benchmarker scheme by a single NSC encoder and a STTC encoder. A bit-based channel interleaver is inserted between the NSC encoder and STTC encoder. Taking into account the bits required for code termination, the number of output bits of the NSC encoder is $(1116 + k_0 L_0)/R_0 = 2244$, which corresponds to 561 16QAM symbols. Again, a 16-state STTC scheme having $N_t = 2$ transmit antennas is employed. After code termination, we have $561 + 1 = 562$ 16QAM symbols or $4(562) = 2248$ bits in a transmission frame at each transmit antenna. The overall coding rate is given by $R = 1116/2248 \approx 0.4964$ and the effective throughput is $\log_2(16)R \approx 1.99$ BPS, both of which are very close to the corresponding values of the STTC-TCM-2NSC-2 scheme. A decoding iteration of the STTC-NSC benchmarker scheme is comprised of a STTC decoding and a NSC decoding step.

We will quantify the decoding complexity of the proposed STTC-TCM-2NSC scheme and that of the benchmarker scheme using the number of decoding trellis states. The total number of decoding trellis states per iteration for the proposed scheme employing 2 NSC decoders having a code memory of $L_1 = L_2 = 3$, TCM having $L_3 = 3$ and STTC having $L_4 = 4$, is given by $S = 2^{L_1} + 2^{L_2} + 2^{L_3} + 2^{L_4} = 40$. By contrast, the total number of decoding trellis states per iteration for the benchmarker scheme having a code memory of $L_0 = 6$ and STTC having $L_4 = 4$, is given by $S = 2^{L_0} + 2^{L_4} = 80$. Therefore, the complexity of the proposed STTC-TCM-2NSC scheme having two iterations is equivalent to that of the benchmarker scheme having a single iteration, which corresponds to 80 decoding states.

11.6.4 Turbo Transceiver Performance Results

In this section we evaluate the performance of the proposed MPEG-4 based audio transceiver schemes using both the achievable Bit Error Ratio (BER) and the attainable Segmental Signal to Noise Ratio (SegSNR).

Figures 11.36 and 11.37 depict the BER versus Signal to Noise Ratio (SNR) per bit, namely E_b/N_0, performance of the 16QAM-based STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes, respectively, when communicating over uncorrelated Rayleigh fading channels. As we can observe from Figures 11.36 and 11.37, the gap between the BER performance of the class-1 and class-2 audio bits is wider for STTC-TCM-2NSC-1 compared to the STTC-TCM-2NSC-2 scheme. More explicitly, the class-1 audio bits of STTC-TCM-2NSC-1 have a higher protection at the cost of a lower throughput compared to the STTC-TCM-2NSC-2 scheme. However, the BER performance of the class-2 audio bits of the STTC-TCM-2NSC-1 arrangement is approximately 0.5 dB poorer than that of STTC-TCM-2NSC-2 at BER=10^{-5}.

Let us now study the audio SegSNR performance of the schemes in Figures 11.38 and 11.39. As we can see from Figure 11.38, the SegSNR performance of STTC-TCM-2NSC-1 is inferior in comparison to that of STTC-TCM-2NSC-2, despite providing a higher protection for the class-1 audio bits. More explicitly, STTC-TCM-2NSC-2 requires E_b/N_0 = 2.5 dB, while STTC-TCM-2NSC-1 requires E_b/N_0 = 3 dB, when having an audio SegSNR in excess of 16 dB after the fourth turbo iteration. Hence the audio SegSNR performance of STTC-TCM-2NSC-1 is 0.5 dB poorer than that of STTC-TCM-2NSC-2 after the fourth iteration. Note that the BER of the class-1 and class-2 audio bits for the corresponding values of E_b/N_0,

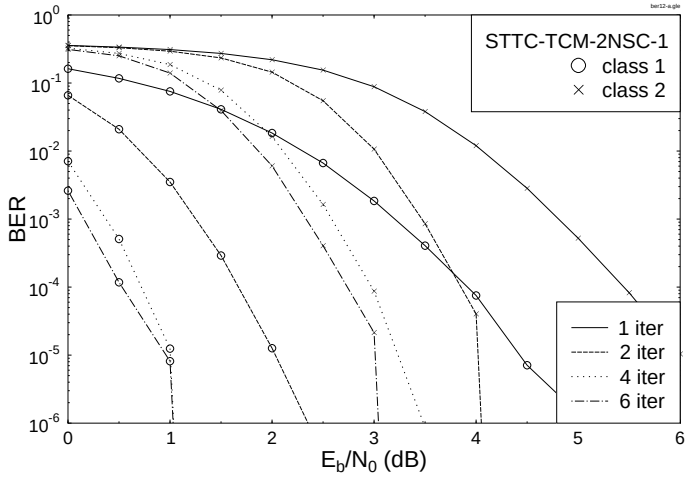

Figure 11.36: BER versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC-1 assisted MPEG-4 audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.77 BPS**.

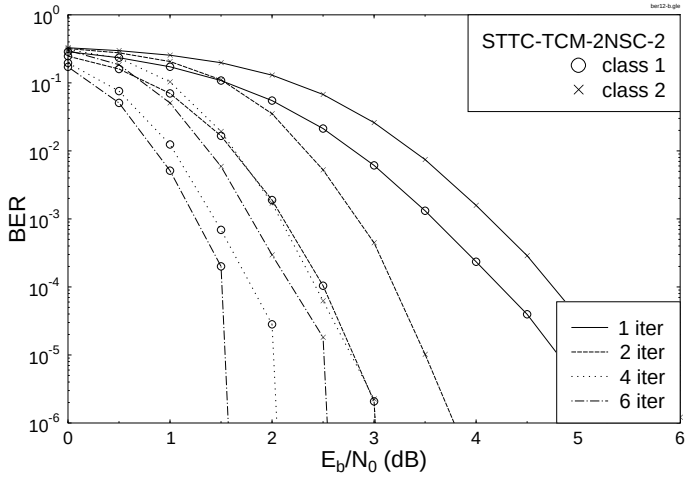

Figure 11.37: BER versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC-2 assisted MPEG-4 audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **2.07 BPS**.

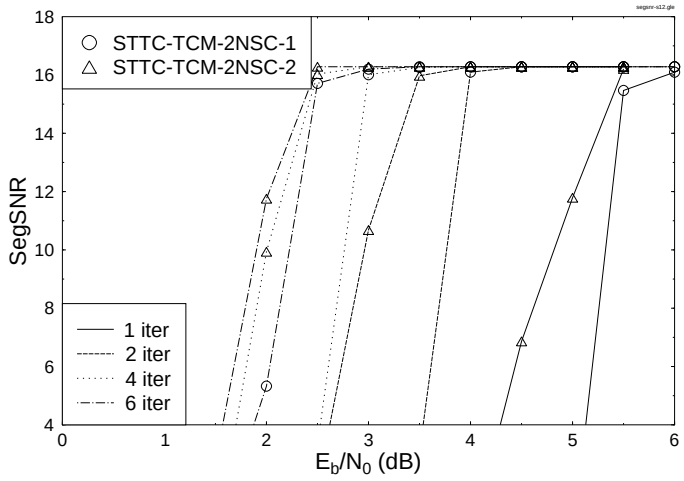

Figure 11.38: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC assisted MPEG-4 audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput of STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 was **1.77** and **2.07 BPS**, respectively.

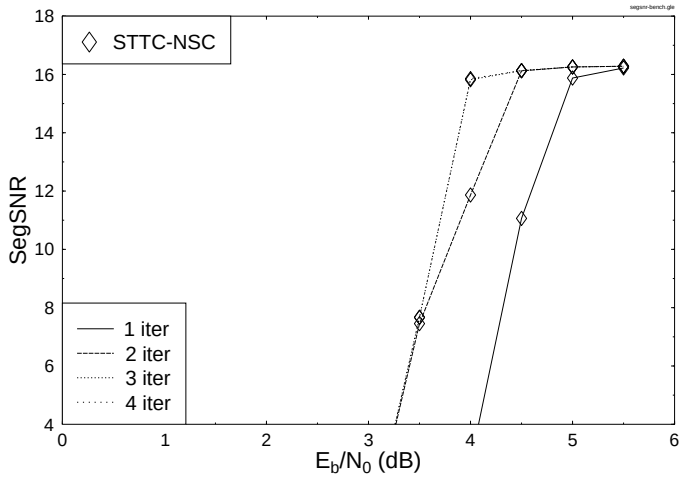

Figure 11.39: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-NSC assisted MPEG-4 audio benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.99 BPS**.

SegSNR and iteration index is less than 10^{-7} and 10^{-4}, respectively, for the two different turbo schemes. After the sixth iteration, the SegSNR performance of both turbo schemes becomes quite similar since the corresponding BER is low. These results demonstrate that the MPEG-4 audio decoder requires a very low BER for both class-1 and class-2 audio bits, when aiming for a SegSNR above 16 dB. In this context it is worth mentioning that Recursive Systematic Convolutional codes (RSCs) [509–511] are capable of achieving a higher iteration gain, but suffer from an error floor. Owing to this reason the SegSNR performance of the schemes employing RSCs instead of NSCs was found to be poorer. The SegSNR results of the turbo schemes employing RSCs instead of NSCs as the outer code were not shown here for reasons of space economy.

Figure 11.39 portrays the SegSNR versus E_b/N_0 performance of the STTC-NSC audio benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. Note that if we reduce the code memory of the NSC constituent code of the STTC-NSC benchmarker arrangement from L_0=6 to 3, the achievable performance becomes poorer, as expected. If we increased L_0 from 6 to 7 (or higher), the decoding complexity would increase significantly, while the attainable best possible performance is only marginally increased. Hence, the STTC-NSC scheme having L_0=6 constitutes a good benchmarker scheme in terms of its performance versus complexity tradeoffs. It is shown in Figures 11.38 and 11.39 that the first iteration based performance of the STTC-NSC benchmarker scheme is better than that of the proposed STTC-TCM-2NSC arrangements. However, at the same decoding complexity of 160 (240) trellis decoding states STTC-TCM-2NSC-2 having 4 (6) iterations performs approximately 2 (1.5) dB better than the STTC-NSC arrangement having 2 (3) iterations.

It is worth mentioning that other joint coding and modulation schemes directly designed for fading channels, such as for example Bit Interleaved Coded Modulation (BICM) [511, 512, 517] were outperformed by the TCM-based scheme, since the STTC arrangement rendered the error statistics more Gaussian-like [518].

11.6.5 MPEG-4 Turbo Transceiver Summary

In conclusion, a jointly optimised audio source-coding, outer unequal protection NSC channel-coding, inner TCM and spatial diversity aided STTC turbo transceiver was proposed for employment in a MPEG-4 wireless audio transceiver. With the aid of two different-rate NSCs the audio bits were protected differently according to their error sensitivity. The employment of TCM improved the bandwidth efficiency of the system and by utilising STTC spatial diversity was attained. The performance of the proposed STTC-TCM-2NSC scheme was enhanced with the advent of an efficient iterative joint decoding structure. The high-compression MPEG-4 audio decoder is sensitive to transmission errors and hence it was found to require a low BER for both classes of audio bits in order to attain a perceptually pleasing, artefact-free audio quality. The proposed twin-class STTC-TCM-2NSC scheme performs approximately 2 dB better in terms of the required E_b/N_0 than the single-class STTC-NSC audio benchmarker.

11.7 Turbo-Detected Space-Time Trellis Coded MPEG-4 Versus AMR-WB Speech Transceivers

N. S. Othman, S. X. Ng and L. Hanzo

11.7.1 Motivation and Background

The MPEG-4 TwinVQ audio codec and the AMR-WB speech codec are investigated in the context of a jointly optimised turbo transceiver capable of providing unequal error protection. The transceiver advocated consists of serially concatenated Space-Time Trellis Coding (STTC), Trellis Coded Modulation (TCM) and two different-rate Non-Systematic Convolutional codes (NSCs) used for unequal error protection. A benchmarker scheme combining STTC and a single-class protection NSC is used for comparison with the proposed scheme. The audio and speech performance of both schemes is evaluated, when communicating over uncorrelated Rayleigh fading channels. We will demonstrate that an E_b/N_0 value of about 2.5 (3.5) dB is required for near-unimpaired audio (speech) transmission, which is about 3.07 (4.2) dB from the capacity of the system.

In recent years, joint source-channel coding (JSCC) has been receiving significant research attention in the context of both delay- and complexity-constrained transmission scenarios. JSCC aims at designing the source codec and channel codec jointly for the sake of achieving the highest possible system performance. As it was argued in [518], this design philosophy does not contradict to the classic Shannonian source and channel coding separation theorem. This is because instead of considering perfectly lossless Shannonian entropy coders for source coding and transmitting their bitstreams over Gaussian channels, we consider low-bitrate lossy audio and speech codecs, as well as Rayleigh-fading channels. Since the bitstreams of the speech and audio encoders are subjected to errors during wireless transmission, it is desirable to provide stronger error protection for the audio bits, which have a substantial effect on the objective or subjective quality of the reconstructed speech or audio signals. Unequal error protection (UEP) is a particular manifestation of JSCC, which offers a mechanism to match the error protection capabilities of channel coding schemes having different error correction capabilities to the differing bit-error sensitivities of the speech or audio bits [519].

Speech services are likely to remain the most important ones in wireless systems. However, there is an increasing demand for high-quality speech transmissions in multimedia applications, such as video-conferencing [514]. Therefore, an expansion of the speech bandwidth from the 300-3400 Hz range to a wider bandwidth of 50-7000 Hz is a key factor in meeting this demand. This is because the low-frequency enhancement ranging from 50 to 200 Hz contributes to the increased naturalness, presence and comfort, whilst the higher-frequency extension spanning from 3400 to 7000 Hz provides a better fricative differentiation and therefore a higher intelligibility. A bandwidth of 50 to 7000 Hz not only improves the intelligibility and naturalness of speech, but also adds an impression of transparent communication and eases speaker recognition. The Adaptive Multi-Rate Wideband (AMR-WB) voice codec has become a 3GPP standard, which provides a superior speech quality [520].

For the sake of supporting high-quality multimedia services over wireless communication channels requires the development of techniques for transmitting not only speech, but also video, music, and data. Therefore, in the field of audio-coding, high-quality, high-

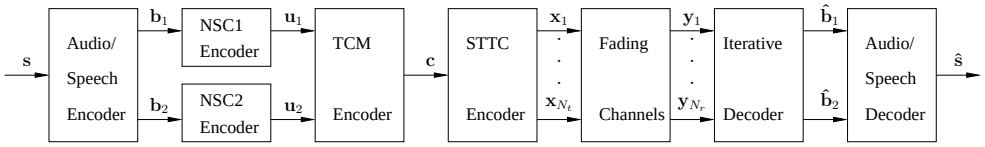

Figure 11.40: Block diagram of the serially concatenated STTC-TCM-2NSC assisted audio/speech scheme. The notations $\mathbf{s}$, $\hat{\mathbf{s}}$, $\mathbf{b}_i$, $\hat{\mathbf{b}}_\mathbf{i}$, $\mathbf{u}_i$, $\mathbf{c}$, $\mathbf{x}_j$ and $\mathbf{y}_k$ denote the vector of the audio/speech source symbol, the estimate of the audio/speech source symbol, the class-i audio/speech bits, the estimates of the class-i audio/speech bits, the encoded bits of class-i NSC encoders, the TCM coded symbols, the STTC coded symbols for transmitter j and the received symbols at receiver k, respectively. Furthermore, N_t and N_r denote the number of transmitters and receivers, respectively. The symbol-based channel interleaver between the STTC and TCM schemes as well as the two bit-based interleavers at the output of NSC encoders are not shown for simplicity. The iterative decoder seen at the right is detailed in Figure 11.43.

compression and highly error-resilient audio-coding algorithms are required. The MPEG-4 Transform-domain Weighted Interleaved Vector Quantization (TwinVQ) scheme is a low-bit-rate audio-coding technique that achieves a high audio quality under error-free transmission conditions at bitrates below 40 kbps [506]. In order to render this codec applicable to wireless systems, which typically exhibit a high bit-error ratio (BER), powerful turbo transceivers are required.

Trellis Coded Modulation (TCM) [510–512] constitutes a bandwidth-efficient joint channel coding and modulation scheme, which was originally designed for transmission over Additive White Gaussian Noise (AWGN) channels. Space-Time Trellis Coding (STTC) [511, 513] is a joint spatial diversity and channel coding technique. STTC may be efficiently employed in an effort to mitigate the effects of Rayleigh fading channels and render them Gaussian-like for the sake of supporting the operation of a TCM code. Recently, a sophisticated unequal-protection turbo transceiver using twin-class convolutional outer coding, as well as joint coding and modulation as inner coding combined with STTC-based spatial diversity scheme was designed for MPEG-4 video telephony in [516, 518]. Specifically, maximal minimum distance Non-Systematic Convolutional codes (NSCs) [509, p. 331] having two different code-rates were used as outer encoders for providing unequal MPEG-4 video protection. Good video quality was attained at a low SNR and medium complexity by the proposed transceiver. By contrast, in this section we study the achievable performance of the AMR-WB and the MPEG-4 TwinVQ speech and audio codecs in conjunction with the sophisticated unequal-protection turbo transceiver of [516, 518].

11.7.2 The AMR-WB Codec'S Error Sensitivity

The synthesis filter's excitation signal in the AMR-WB codec is based on the Algebraic Code Excited Linear Predictor (ACELP) algorithm, supporting nine different speech codec modes having bitrates of 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85 and 6.6 kbps [520]. Like most ACELP-based algorithms, the AMR-WB codec interprets 20 ms segments of speech as the output of a linear synthesis filter synthesized from an appropriate excitation signal. The task of the encoder is to optimise the filter as well as the excitation signal and

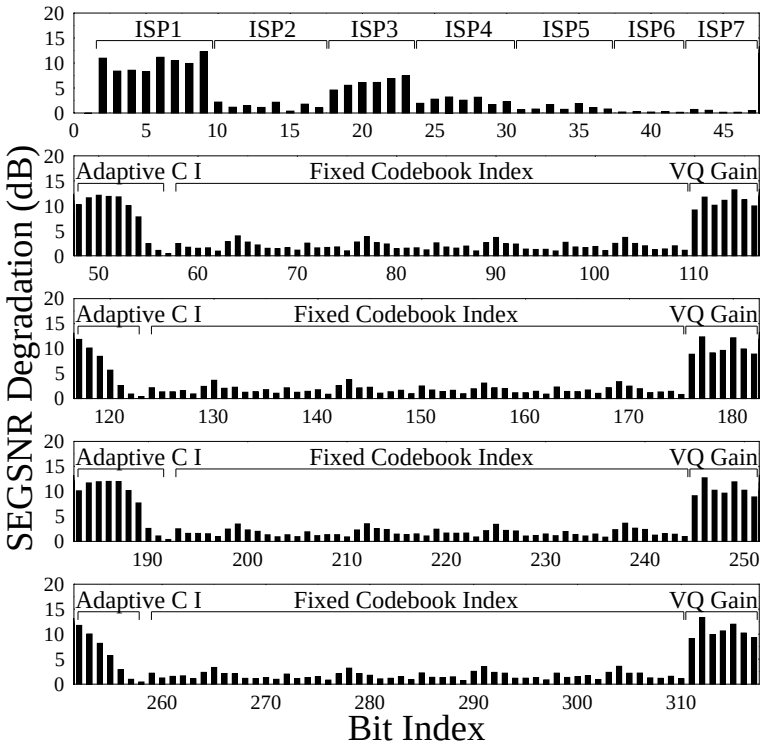

Figure 11.41: SegSNR degradations versus bit index due to inflicting 100% Bit Error Rate (BER) in the 317-bit, 20 ms AMR-WB frame

then represent both as efficiently as possible with the aid of a frame of binary bits. At the decoder, the encoded bit-based speech description is used to synthesize the speech signal by inputting the excitation signal to the synthesis filter, thereby generating the speech segment. Again, each AMR-WB frame represents 20 ms of speech, producing 317 bits at a bitrate of 15.85 kbps. The codec parameters that are transmitted over the noisy channel include the so-called imittance spectrum pairs (ISPs), the adaptive codebook delay (pitch delay), the algebraic codebook excitation index, and the jointly vector quantized, pitch gains as well as algebraic codebook gains.

Most source coded bitstreams contain certain bits that are more sensitive to transmission errors than others. A common approach used for quantifying the sensitivity of a given bit is to consistently invert this bit in every speech or audio frame and evaluate the associated Segmental SNR (SegSNR) degradation [515]. The SegSNR degradation is computed by subtracting from the SegSNR recorded under error-free conditions the corresponding value when there are channel-induced bit-errors.

The error sensitivity of the various encoded bits in the AMR-WB codec determined in this way is shown in Figure 2. The results are based on samples taken from the EBU SQAM (Sound Quality Assessment Material) CD, sampled at 16 kHz and encoded at 15.85 kbps. It can be observed that the bits representing the ISPs, the adaptive codebook delay, the algebraic codebook index and the vector quantized gain are fairly error sensitive. The least sensitive

bits are related to the fixed codebook's excitation pulse positions, as shown in Figure 11.41. This is because, when one of the fixed codebook index bits is corrupted, the codebook entry selected at the decoder will differ from that used in the encoder only in the position of one of the non-zero excitation pulses. Therefore the corrupted excitation codebook entry will be similar to the original one. Hence, the algebraic codebook structure used in the AMR-WB codec is quite robust to channel errors.

11.7.3 The MPEG-4 TwinVQ Codec'S Error Sensitivity

The MPEG-4 TwinVQ scheme is a transform coder that uses the modified discrete cosine transformation (MDCT) [506] for transforming the input signal into the frequency-domain transform coefficients. The input signal is classified into one of three modes, each associated with a different transform window size, namely a long, medium or short window, catering for different input signal characteristics. The MDCT coefficients are normalized by the spectral envelope information obtained through the Linear Predictive Coding (LPC) analysis of the signal. Then the normalized coefficients are interleaved and divided into sub-vectors by using the so-called interleave and division technique of [506], and all sub-vectors are encoded separately by the VQ modules.

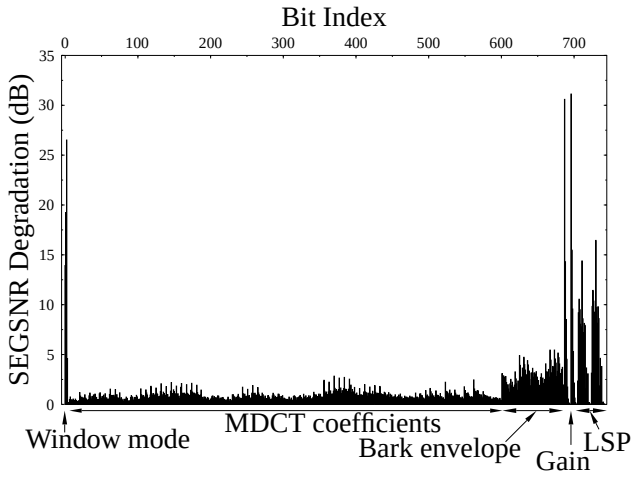

Figure 11.42: SegSNR degradations due to inflicting a 100% BER in the 743-bit, 23.22 ms MPEG-4 TwinVQ frame

Similarly, bit error sensitivity investigations were performed in the same way, as described in the previous section. Figure 11.42 shows the error sensitivity of the various bits of the MPEG-4 TwinVQ codec for a bitrate of 32 kbps. The results provided are based on a 60 seconds long excerpt of Mozart's "Clarinet Concerto (2nd movement - Adagio)". This stereo audio file was sampled at 44.1 kHz and again, encoded at 32 kbps. Since the analysis frame length is 23.22 ms, which corresponds to 1024 audio input samples, there are 743 encoded bits in each frame. This figure shows that the bits representing the gain factors, the Line Spectral Frequency (LSF) parameters, and the Bark-envelope are more sensitive to channel errors, compared to the bits representing the MDCT coefficients. The bits signalling the

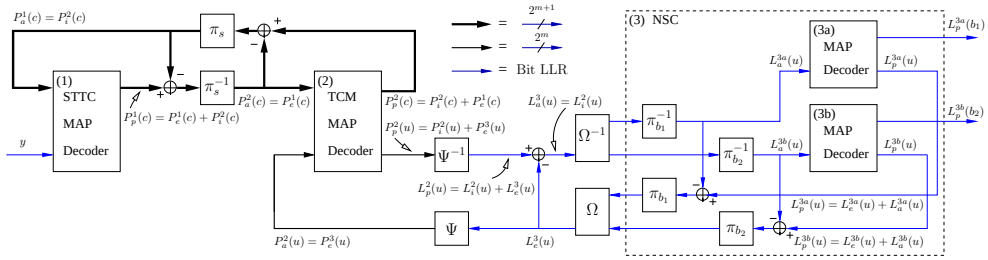

Figure 11.43: Block diagram of the STTC-TCM-2NSC turbo detection scheme seen at the right of Figure 11.40. The notations $\pi_{(s,b_i)}$ and $\pi^{-1}_{(s,b_i)}$ denote the interleaver and deinterleaver, while the subscript s denotes the symbol-based interleaver of TCM and the subscript b_i denotes the bit-based interleaver for class-i NSC. Furthermore, Ψ and Ψ^{-1} denote LLR-to-symbol probability and symbol probability-to-LLR conversion, while Ω and Ω^{-1} denote the parallel-to-serial and serial-to-parallel converter, respectively. The notation m denotes the number of information bits per TCM coded symbol [516] ©IEE, 2004, Hanzo.

window mode used are also very sensitive to transmission errors and hence have to be well protected. The proportion of sensitive bits was only about 10%. This robustness is deemed to be a benefit of the weighted vector-quantization procedure which uses a fixed-length coding structure as opposed to using an error-sensitive variable-length structure, where transmission errors would result in a loss of synchronisation.

11.7.4 The Turbo Transceiver

Once the bit error sensitivity of the audio/speech codecs was determined, the bits of the AMR-WB and the MPEG-4 TwinVQ codec are protected according to their relative importance. Figure 11.40 shows the schematic of the serially concatenated STTC-TCM-2NSC turbo scheme using a STTC and a TCM scheme as well as two different-rate NSCs as its constituent codes. Let us denote the turbo scheme using the AMR-WB codec as STTC-TCM-2NSC-AMR-WB, whilst STTC-TCM-2NSC-TVQ refers to the turbo scheme using the MPEG-4 TwinVQ as the source codec. For comparison, both schemes protect 25% of the most sensitive bits in class-1 using an NSC code rate of $R_1 = k_1/n_1 = 1/2$. By contrast, the remaining 75% of the bits in class-2 are protected by an NSC scheme having a rate of $R_2 = k_2/n_2 = 3/4$. The code memory of the class-1 and class-2 encoders is $L_1 = 3$ and $L_2 = 3$, respectively. The class-1 and class-2 NSC coded bit sequences are interleaved by two separate bit interleavers, before they are fed to the rate-$R_3 = 3/4$ TCM scheme [510–512] having a code memory of $L_3 = 3$. Code termination was employed for the NSCs, as well as for the TCM [510–512] and STTC codecs [511,513]. The TCM symbol sequence is then symbol-interleaved and fed to the STTC encoder as seen in Figure 11.43. We invoke a 16-state STTC scheme having a code memory of $L_4 = 4$ and $N_t = 2$ transmit antennas, employing $M =$ 16-level Quadrature Amplitude Modulation (16QAM) [512]. The STTC scheme employing $N_t = 2$ requires a single 16QAM-based termination symbol. In the STTC-TCM-2NSC-AMR-WB scheme the 25% of the bits that are classified into class-1 includes 23 header bits, which gives a total of 340 NSC1-encoded bits. In the ITU stream format [521], the header bits of each frame include the frame types and the window-mode

used.

Hence, the overall coding rate of the STTC-TCM-2NSC-AMR-WB scheme becomes $R_{AMRWB} = 340/720 \approx 0.4722$. By contrast, the overall coding rate of the STTC-TCM-2NSC-TVQ scheme is $R_{TVQ} = 744/1528 \approx 0.4869$. The effective throughput of the STTC-TCM-2NSC-AMR-WB and STTC-TCM-2NSC-TVQ schemes is $log_2(M) \cdot R_{AMRWB} \approx 1.89$ Bits Per Symbol (BPS) and $log_2(M) \cdot R_{TVQ} \approx 1.95$ BPS, respectively.

At the receiver, we employ $N_r = 2$ receive antennas and the received signals are fed to the iterative decoders for the sake of estimating the audio bit sequences in both class-1 and class-2, as seen in Figure 11.40. The STTC-TCM-2NSC scheme's turbo decoder structure is illustrated in Figure 11.43, where there are four constituent decoders, each labelled with a round-bracketed index. The Maximum A-Posteriori (MAP) algorithm [511] operating in the logarithmic-domain is employed by the STTC and TCM schemes as well as by the two NSC decoders, respectively. The iterative turbo-detection scheme shown in Figure 11.43 enables an efficient information exchange between the STTC, TCM and NSCs constituent codes for the sake of achieving spatial diversity gain, coding gain, unequal error protection and a near-channel-capacity performance. The information exchange mechanism between each constituent decoders is detailed in [516].

For the sake of benchmarking both audio schemes advocated, we created a powerful benchmark scheme for each of them by replacing the TCM and NSC encoders of Figure 11.40 by a single-class NSC codec having a coding rate of $R_0 = k_0/n_0 = 1/2$ and a code memory of $L_0 = 6$. Note that if we reduce the code memory of the NSC constituent code of the STTC-NSC benchmarker arrangement from L_0=6 to 3, the achievable performance becomes poorer, as expected. If we increased L_0 from 6 to 7 (or higher), the decoding complexity would double, while the attainable performance is only marginally increased. Hence, the STTC-NSC scheme having L_0=6 constitutes a good benchmarker scheme in terms of its performance versus complexity tradeoffs. We will refer to this benchmarker scheme as the STTC-NSC-TVQ and the STTC-NSC-AMR-WB arrangement designed for the audio and the speech transceiver, respectively. Again, all audio and speech bits are equally protected in the benchmarker scheme by a single NSC encoder and a STTC encoder. A bit-based channel interleaver is inserted between the NSC encoder and STTC encoder. Taking into account the bits required for code termination, the number of output bits of the NSC encoder of the STTC-NSC-TVQ benchmarker scheme is $(744 + k_0L_0)/R_0 = 1500$, which corresponds to 375 16QAM symbols. By contrast, in the STTC-NSC-AMR-WB scheme the number of output bits after taking into account the bits required for code termination becomes $(340 + k_0L_0)/R_0 = 692$, which corresponds to 173 16QAM symbols. Again, a 16-state STTC scheme having $N_t = 2$ transmit antennas is employed. After code termination, we have $375 + 1 = 376$ 16QAM symbols or $4(376) = 1504$ bits in a transmission frame at each transmit antenna for the STTC-NSC-TVQ. The overall coding rate is given by $R_{TVQ-b} = 744/1504 \approx 0.4947$ and the effective throughput is $\log_2(16)R_{TVQ-b} \approx 1.98$ BPS, both of which are very close to the corresponding values of the STTC-TCM-2NSC-TVQ scheme. Similary, for the STTC-NSC-AMR-WB scheme, after code termination, we have $173 + 1 = 174$ 16QAM symbols or $4(174) = 696$ bits in a transmission frame at each transmit antenna. This gives the overall coding rate as $R_{AMRWB-b} = 340/696 \approx 0.4885$ and the effective throughput becomes $\log_2(16)R_{AMRWB-b} \approx 1.95$ BPS. Again, both of the values are close to the corresponding values of the STTC-TCM-2NSC-AMR-WB scheme. A decoding iteration of each of the STTC-NSC benchmarker schemes is comprised of a STTC

decoding and a NSC decoding step.

We will quantify the decoding complexity of the proposed STTC-TCM-2NSC schemes and that of its corresponding benchmarker schemes using the number of decoding trellis states. The total number of decoding trellis states per iteration of the proposed scheme employing 2 NSC decoders having a code memory of $L_1 = L_2 = 3$, using the TCM scheme having $L_3 = 3$ and the STTC arrangement having $L_4 = 4$, becomes $S = 2^{L_1} + 2^{L_2} + 2^{L_3} + 2^{L_4} = 40$. By contrast, the total number of decoding trellis states per iteration for the benchmarker scheme having a code memory of $L_0 = 6$ and for the STTC having $L_4 = 4$ is given by $S = 2^{L_0} + 2^{L_4} = 80$. Therefore, the complexity of the proposed STTC-TCM-2NSC scheme having two iterations is equivalent to that of the benchmarker scheme having a single iteration, which corresponds to 80 decoding states.

11.7.5 Performance Results

In this section we comparatively study the performance of the audio and speech transceiver using the Segmental Signal to Noise Ratio (SegSNR) metric.

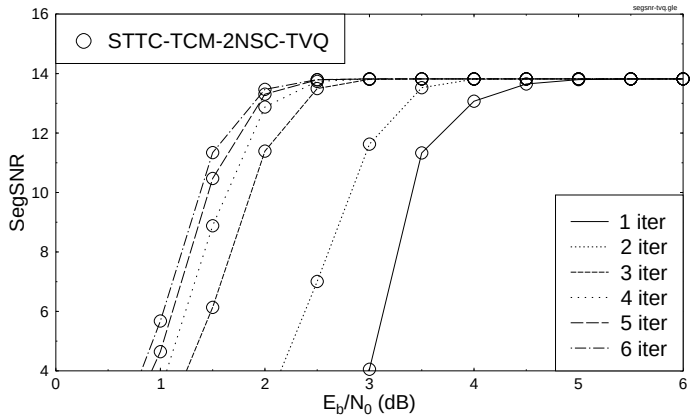

Figure 11.44: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC assisted MPEG-4 TwinVQ audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.95 BPS**.

Figures 11.44 and 11.45 depict the audio SegSNR performance of the STTC-TCM-2NSC-TVQ and that of its corresponding STTC-NSC-TVQ benchmarker schemes, respectively, when communicating over uncorrelated Rayleigh fading channels. It can be seen from Figures 11.44 and 11.45 that the non-iterative single-detection based performance of the STTC-NSC-TVQ benchmarker scheme is better than that of the STTC-TCM-2NSC assisted MPEG-4 TwinVQ audio scheme. However, at the same decoding complexity quantified in terms of the number of trellis decoding states the STTC-TCM-2NSC-TVQ arrangement performs approximately 0.5 dB better in terms of the required channel E_b/N_0 value than the STTC-NSC-TVQ benchmarker scheme, both exhibiting a SegSNR of 13.8 dB. For example, at the decoding complexity of 160 trellis decoding states, this corresponds to the STTC-TCM-2NSC-TVQ scheme's 4th iteration, whilst in the STTC-NSC-TVQ scheme this corresponds to the 2nd iteration. Therefore, we observe in Figures 11.44 and 11.45 that the STTC-TCM-

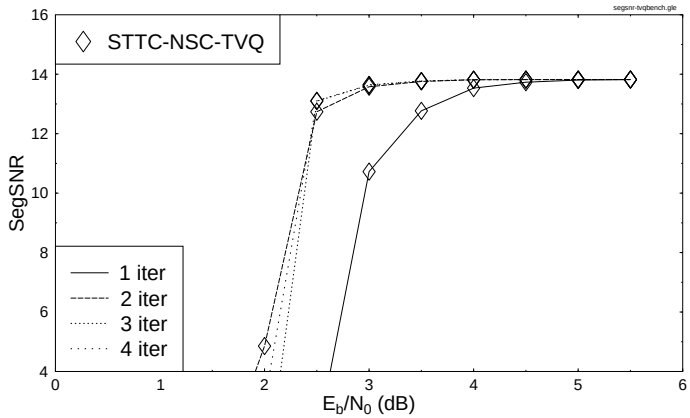

Figure 11.45: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-NSC assisted MPEG-4 TwinVQ audio benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.98 BPS**.

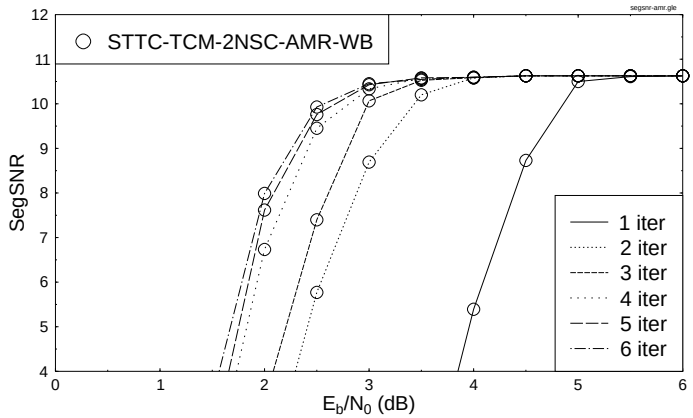

Figure 11.46: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC assisted AMR-WB speech scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.89 BPS**.

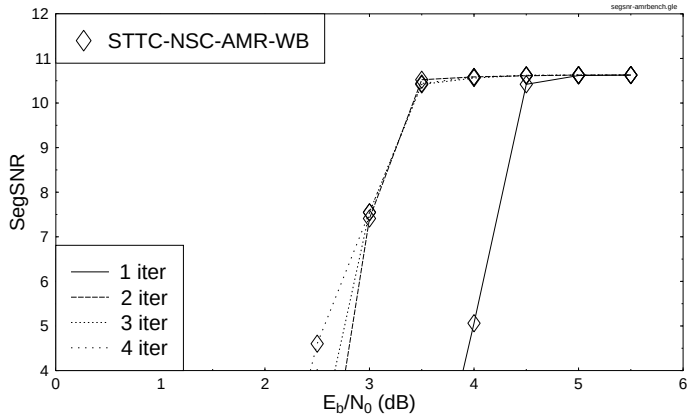

Figure 11.47: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-NSC assisted AMR-WB speech benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.95 BPS**.

2NSC-TVQ arrangement performs by 0.5 dB better in terms of the required channel E_b/N_0 value than its corresponding benchmarker scheme.

Similarly, it can be observed from Figures 11.46 and 11.47 that at the decoding complexity of 160 trellis decoding states the STTC-TCM-2NSC-AMR-WB arrangement performs 0.5 dB better in terms of the required channel E_b/N_0 value than the STTC-NSC-AMR-WB scheme, when targetting a SegSNR of 10.6 dB. By comparing Figures 11.44 and 11.46, we observe that the SegSNR performance of the STTC-TCM-2NSC-AMR-WB scheme is inferior in comparison to that of STTC-TCM-2NSC-TVQ.

More explicitly, the STTC-TCM-2NSC-TVQ system requires an E_b/N_0 value of 2.5 dB, while the STTC-TCM-2NSC-AMR-WB arrangement necessitates $E_b/N_0 = 3.0$ dB, when having their respective maximum attainable average SegSNRs. The maximum attainable average SegSNRs for STTC-TCM-2NSC-TVQ and STTC-TCM-2NSC-AMR-WB are 13.8 dB and 10.6 dB, respectively.

This discrepancy is due to the reason that both schemes map the most sensitive 25% of the encoded bits to class-1. By contrast, based on the bit error sensitivity study of the MPEG-4 TwinVQ codec outlined in Section 3, only 10% of the MPEG-4 TwinVQ encoded bits were found to be gravely error sensitive. Therefore, the 25% class-1 bits of the MPEG-4 TwinVQ also includes some bits, which were found to be only moderately sensitive to channel errors. However, in the case of the AMR-WB codec all the bits of the 25%-partition were found to be quite sensitive to channel errors. Furthermore, the frame length of the STTC-TCM-2NSC-TVQ scheme is longer than that of the STTC-TCM-2NSC-AMR-WB arrangement and hence benefits from a higher coding gain.

It is worth mentioning that the channel capacity for the system employing the full-diversity STTC scheme with the aid of $N_t = 2$ transmit antennas and $N_r = 2$ receive antennas is -0.57 dB and -0.70 dB for the throughputs of 1.95 BPS and 1.89 BPS, respectively, when communicating over uncorrelated Rayleigh fading channels [522].

11.7.6 AMR-WB and MPEG-4 TwinVQ Turbo Transceiver Summary

In this section we comparatively studied the performance of the MPEG-4 TwinVQ and AMR-WB audio/speech codecs combined with a jointly optimised source-coding, outer unequal protection NSC channel-coding, inner TCM and spatial diversity aided STTC turbo transceiver. The audio bits were protected differently according to their error sensitivity with the aid of two different-rate NSCs . The employment of TCM improved the bandwidth efficiency of the system and by utilising STTC spatial diversity was attained. The performance of the STTC-TCM-2NSC scheme was enhanced with the advent of an efficient iterative joint decoding structure. Both proposed twin-class STTC-TCM-2NSC schemes perform approximately 0.5 dB better in terms of the required E_b/N_0 than the corresponding single-class STTC-NSC audio benchmarker schemes. This relatively modest advantage of the twin-class protected transceiver was a consequence of having a rather limited turbo-interleaver length. In the longer interleaver of the videohphone system of [516, 518] an approximately 2 dB E_b/N_0 gain was achieved. For a longer-delay non-realtime audio streaming scheme a simliar performance would be achieved to that of [516]. Our future work will further improve the achievable audio performance using the soft speech-bit decoding technique of [523].

11.8 Chapter Summary

In this chapter the MPEG-4 Audio standard was discussed in detail. The MPEG-4 audio standard is constituted by a toolbox of different coding algorithms, designed for coding both speech and music signals in the range spanning from very low bit rates, such as 2 kbit/s to rates as high as 64 kbit/s. In Section 11.2, the important milestones in the field of audio coding were described and summarised in Figure 11.2. Specifically, four key technologies, namely perceptual coding, frequency domain coding, the window switching strategy and the dynamic bit allocation technique were fundamentally important in the advancement of audio coding. The MPEG-2 AAC codec [40], as described in Section 11.2.1 forms a core part of the MPEG-4 audio codec. Various tools that can be used for processing the transform coefficients in order to achieve an improved coding efficiency were highlighted in Sections 11.2.2 to 11.2.5. The AAC quantization procedure was discussed in Section 11.2.6, while two other tools provided for encoding the transform coefficients, namely the BSAC and TWIVQ techniques were detailed in Sections 11.2.8 and 11.2.9, respectively. More specifically, the BSAC coding technique provides finely-grained bitstream scalability, in order to further reduce the redundancy inherent in the quantized spectrum of the audio signal generated by the MPEG-4 codec. The TWINVQ codec [463] described in Section 11.2.9 was found to be capable of encoding both speech and music signals, which provides an attractive option for low bit rate audio coding.

In Section 11.3, which was dedicated to speech coding tools, the HVXC and CELP codecs were discussed. The HVXC codec was employed for encoding speech signals in the bit rate range spanning from 2 to 4 kbit/s, while the CELP codec is used at bit rates between 4 and 24 kbit/s, with the additional capability of encoding speech signals at the sampling rates of 8 and 16 kHz.

In Section 11.5 turbo coded and space-time coded adaptive as well as fixed modulation based OFDM assisted MPEG-4 audio systems have been investigated. The transmission parameters have been partially harmonised with the UMTS TDD mode [491], which provides an

attractive system design framework. More specifically, we employed the MPEG-4 TWINVQ codec at the bit rates of 16, 32 and 64 kbit/s. We found that by employing space-time coding, the channel quality variations have been significantly reduced and no additional benefits could be gained by employing adaptive modulation. However, adaptive modulation was found beneficial when it was employed in a low-complexity one-transmitter, one-receiver scenario when high channel quality variations were observed. The space-time coded, two-transmitter, one-receiver configuration was shown to outperform the conventional one-transmitter, one-receiver scheme by about 4 dB in channel SNR terms over the highly dispersive COST207 TU channel.

In Section 11.7.6 we comparatively studied the performance of the MPEG-4 TwinVQ and AMR-WB audio/speech codecs combined with a jointly optimised source-coding, outer unequal protection NSC channel-coding, inner TCM and spatial diversity aided STTC turbo transceiver. The employment of TCM provided further error protection without expanding the bandwidth of the system and by utilising STTC spatial diversity was attained, which rendered the error statistics experienced pseudo-random, as required by the TCM scheme, since it was designed for Gaussian channels inflicting randomly dispersed channel errors. Finally, the performance of the STTC-TCM-2NSC scheme was enhanced with the advent of an efficient iterative joint decoding structure. Both proposed twin-class STTC-TCM-2NSC schemes perform approximately 0.5 dB better in terms of the required E_b/N_0 than the corresponding single-class STTC-NSC audio benchmarker schemes. This relatively modest advantage of the twin-class protected transceiver was a consequence of having a rather limited turbo-interleaver length imposed by the limited tolertable audio delay. In the longer interleaver of the less delay-limited videophone system of [516, 518] an approximately 2 dB E_b/N_0 gain was achieved. For a longer-delay non-realtime audio streaming scheme a similiar performance would be achieved to that of [516].

Part IV

Very Low Rate Coding and Transmission

Bibliography

[1] R. Cox and P. Kroon, "Low bit-rate speech coders for multimedia communications," *IEEE Communications Magazine*, pp. 34–41, December 1996.

[2] R. Cox, "Speech coding and synthesis," in *Speech coding standards* (W. Kleijn and K. Paliwal, eds.), ch. 2, pp. 49–78, Netherlands: Elsevier, 1995.

[3] R. Steele, *Delta modulation systems*. London, UK: Pentech Press, 1975.

[4] K. Cattermole, *Principles of Pulse Code Modulation*. London, UK: Hiffe Books, 1969.

[5] J. Markel and A. Gray Jr., *Linear Prediction of Speech*. New York, USA: Springer-Verlag, 1976.

[6] L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1978.

[7] B. Lindblom and S. Ohman, *Frontiers of Speech Communication Research*. New York, USA: Academic Press, 1979.

[8] J. Tobias, ed., *Foundations of Modern Auditory Theory*. New York, USA: Academic Press, 1970. ISBN: 0126919011.

[9] B. Atal and J. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bit rates," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'82* [636], pp. 614–617.

[10] N. Jayant and P. Noll, *Digital Coding of Waveforms, Principles and Applications to Speech and Video*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.

[11] P. Kroon, E. Deprettere, and R. Sluyter, "Regular pulse excitation — a novel approach to effective efficient multipulse coding of speech," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, pp. 1054–1063, October 1986.

[12] P. Vary and R. Sluyter, "MATS-D speech codec: Regular-pulse excitation LPC," in *Proceedings of the Nordic Seminar on Digital Land Mobile Radio Communications (DMRII)*, (Stockholm, Sweden), pp. 257–261, October 1986.

[13] P. Vary and R. Hoffmann, "Sprachcodec für das europäische Funkfernsprechnetz," *Frequenz 42 (1988) 2/3*, pp. 85–93, 1988.

[14] W. Hess, *Pitch determination of speech signals: algorithms and devices*. Berlin: Springer Verlag, 1983.

[15] G. Gordos and G. Takacs, *Digital Speech Processing (Digitalis Beszed Feldolgozas)*. Budapest, Hungary: Technical Publishers (Muszaki Kiado), 1983. in Hungarian.

[16] M. Schroeder and B. Atal, "Code excited linear prediction (CELP): High-quality speech at very low bit rates," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'85*, (Tampa, Florida, USA), pp. 937–940, IEEE, 26–29 March 1985.

[17] D. O'Shaughnessy, *Speech Communication: Human and Machine*. Addison-Wesley, 1987. ISBN: 0780334493.

[18] P. Papamichalis, *Practical Approaches to Speech Coding*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1987.

[19] J. Deller, J. Proakis, and J. Hansen, *Discrete-time processing of speech signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1987.

[20] P. Lieberman and S. Blumstein, *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press, 1988.

[21] S. Quackenbush, T. Barnwell III, and M. Clements, *Objective measures of speech quality*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.

[22] S. Furui, *Digital Speech Processing, Synthesis and Recognition*. Marcel Dekker, 1989.

[23] R. Steele, C.-E. Sundberg, and W. Wong, "Transmission of log-PCM via QAM over Gaussian and Rayleigh fading channels," *IEE Proceedings*, vol. 134, Pt. F, pp. 539–556, October 1987.

[24] R. Steele, C.-E. Sundberg, and W. Wong, "Transmission errors in companded PCM over Gaussian and Rayleigh fading channels," *AT&T Bell Laboratories Technical Journal*, pp. 995–990, July–August 1984.

[25] C.-E. Sundberg, W. Wong, and R. Steele, "Weighting strategies for companded PCM transmitted over Rayleigh fading and Gaussian channels," *AT&T Bell Laboratories Technical Journal*, vol. 63, pp. 587–626, April 1984.

[26] W. Wong, R. Steele, and C.-E. Sundberg, "Soft decision demodulation to reduce the effect of transmission errors in logarithmic PCM transmitted over Rayleigh fading channels," *AT&T Bell Laboratories Technical Journal*, vol. 63, pp. 2193–2213, December 1984.

[27] J. Hagenauer, "Source-controlled channel decoding," *IEEE Transactions on Communications*, vol. 43, pp. 2449–2457, September 1995.

[28] "GSM 06.90: Digital cellular telecommunications system (Phase 2+)." Adaptive Multi-Rate (AMR) speech transcoding, version 7.0.0, Release 1998.

[29] S. Bruhn, E. Ekudden, and K. Hellwig, "Adaptive Multi-Rate: A new speech service for GSM and beyond," in *Proceedings of 3rd ITG Conference on Source and Channel Coding*, (Technical Univ. Munich, Germany), pp. 319–324, 17th-19th, January 2000.

[30] L. Chiariglione, "MPEG: a technological basis for multimedia application," *IEEE Multimedia*, vol. 2, pp. 85–89, Spring 1995.

[31] L. Chiariglione, "The development of an integrated audiovisual coding standard: MPEG," *Proceedings of the IEEE*, vol. 83, pp. 151–157, Feb 1995.

[32] L. Chiariglione, "MPEG and multimedia communications," *IEEE Transactions on Circuits and Systems for Video Technology,*, vol. 1, pp. 5–18, Feb 1997.

[33] "Information technology - Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s - Part 3: Audio, IS11172-3 ." ISO/IEC JTC1/SC29/WG11/ MPEG-1, 1992.

[34] D. Pan, "A Tutorial on MPEG/Audio Compression," *IEEE Multimedia*, vol. 2, no. 2, pp. 60–74, Summer 1995.

[35] K. Brandenburg and G. Stoll, "ISO-MPEG1 Audio: A Generic Standard for Coding of High- Quality Digital Audio," *Journal of Audio Engineering Society*, vol. 42, pp. 780–792, Oct 1994.

[36] S. Shlien, "Guide to MPEG1 Audio Standard," *IEEE Transactions on Braodcasting*, vol. 40, pp. 206–218, Dec 1994.

[37] P. Noll, "MPEG Digital Audio Coding," *IEEE Signal Processing Magazine*, vol. 14, pp. 59–81, Sept 1997.

[38] K. Brandenburg and M. Bosi, "Overview of MPEG Audio: Current and Future Standards for Low-Bit-Rate Audio Coding," *Journal of Audio Engineering Society*, vol. 45, pp. 4–21, Jan/Feb 1997.

[39] "Information technology - Generic coding of moving pictures and associated audio - Part 3: Audio, IS13818-3 ." ISO/IEC JTC1/SC29/WG11/ MPEG-2, 1994.

[40] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding," *Journal of Audio Engineering Society*, vol. 45, pp. 789–814, Oct 1997.

[41] ISO/IEC JTC1/SC29/WG11/N2203, MPEG-4 Audio Version 1 Final Committee Draft 14496-3, `http://www.tnt.uni-hannover.de/project/mpeg/audio/documents/`, March 1998.

[42] R. Koenen, "MPEG-4 Overview." `http://www.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm`.

[43] S. R. Quackenbush, "Coding of Natural Audio in MPEG-4," in *Proceedings of ICASSP, Seattle, Washington, USA*, vol. 6, pp. 3797–3800, 12-15th, May 1998.

[44] K. Brandenburg, O. Kunz, and A. Sugiyama, "MPEG-4 Natural Audio Coding," *Signal Processing:Image Communication*, vol. 15, no. 4, pp. 423–444, 2000.

[45] L. Contin, B. Edler, D. Meares, and P. Schreiner, "Tests on MPEG-4 audio codec proposals," *Signal Processing:Image Communication*, vol. 9, pp. 327–342, May 1997.

[46] ISO/IEC JTC1/SC29/WG11/N2203, MPEG-4 Audio Version 2 Final Committee Draft 14496-3 AMD1, `http://www.tnt.uni-hannover.de/project/mpeg/audio/documents/w2803.html`, July 1999.

[47] E. D. Scheirer, "The MPEG-4 Structured Audio Standard," in *Proceedings of ICASSP, Seattle, Washington, USA*, vol. 6, pp. 3801–3804, 12-15th, May 1998.

[48] B. Vercoe, W. Gardner, and E. Scheirer, "Structured Audio: Creation, transmission and rendering of parametric sound representation," *Proceedings of the IEEE*, vol. 86, pp. 922–940, May 1998.

[49] B. Edler, "Speech Coding in MPEG-4," *International Journal of Speech Technology*, vol. 2, pp. 289–303, May 1999.

[50] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1451–1458, Oct 1998.

[51] L. Hanzo, W. T. Webb, and T. Keller, *Single and Multicarrier Quadrature Amplitude Modulation : Principles and Applications for Personal Communications, WLANs and Broadcasting (2nd Ed.)*. John Wiley-IEEE Press, April 2000.

[52] B. Atal, V. Cuperman, and A. Gersho, eds., *Advances in Speech Coding*. Dordrecht: Kluwer Academic Publishers, January 1991. ISBN: 0792390911.

[53] A. Ince, ed., *Digital Speech Processing: Speech Coding, Synthesis and Recognition*. Dordrecht: Kluwer Academic Publishers, 1992.

[54] J. Anderson and S. Mohan, *Source and Channel Coding — An Algorithmic Approach*. Dordrecht: Kluwer Academic Publishers, 1993.

[55] A. Kondoz, *Digital Speech: Coding for low bit rate communications systems*. New York, USA: John Wiley, 1994.

[56] W. Kleijn and K. Paliwal, eds., *Speech Coding and Synthesis*. The Netherlands: Elsevier Science, 1995.

[57] C. Shannon, *Mathematical Theory of Communication*. University of Illinois Press, 1963.

[58] J. Hagenauer, "Quellengesteuerte kanalcodierung fuer sprach- und tonuebertragung im mobilfunk," *Aachener Kolloquium Signaltheorie*, pp. 67–76, 23–25 March 1994.

[59] A. Viterbi, "Wireless digital communications: A view based on three lessons learned," *IEEE Communications Magazine*, pp. 33–36, September 1991.

[60] S. Lloyd, "Least squares quantisation in PCM," *Institute of Mathematical Statistics Meeting, Atlantic City, NJ, USA*, September 1957.

[61] S. Lloyd, "Least squares quantisation in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–136, 1982.

[62] J. Max, "Quantising for minimum distortion," *IRE Transactions on Information Theory*, vol. 6, pp. 7–12, 1960.

[63] W. Bennett, "Spectra of quantised signals," *Bell System Technical Journal*, pp. 446–472, July 1946.

[64] H. Holtzwarth, "Pulse code modulation und ihre verzerrung bei logarithmischer quanteilung," *Archiv der Elektrischen Uebertragung*, pp. 227–285, January 1949.

[65] P. Panter and W. Dite, "Quantisation distortion in pulse code modulation with nonuniform spacing of levels," *Proceedings of the IRE*, pp. 44–48, January 1951.

[66] B. Smith, "Instantaneous companding of quantised signals," *Bell System Technical Journal*, pp. 653–709, 1957.

[67] P. Noll and R. Zelinski, "A contribution to the quantisation of memoryless model sources," *Technical Report, Heinrich Heinz Institute, Berlin*, 1974. (in German).

[68] M. Paez and T. Glisson, "Minimum mean squared error quantisation in speech PCM and DPCM systems," *IEEE Transactions on Communications*, pp. 225–230, April 1972.

[69] A. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[70] R. Salami, *Robust Low Bit Rate Analysis-by-Synthesis Predictive Speech Coding*. PhD thesis, University of Southampton, UK, 1990.

[71] R. Salami, L. Hanzo, R. Steele, K. Wong, and I. Wassell, "Speech coding," in Steele and Hanzo [575], ch. 3, pp. 186–346.

[72] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[73] W. Webb, "Sizing up the microcell for mobile radio communications," *IEE Electronics and communications Journal*, vol. 5, pp. 133–140, June 1993.

[74] K. Wong and L. Hanzo, "Channel coding," in Steele and Hanzo [575], ch. 4, pp. 347–488.

[75] A. Jennings, *Matrix Computation for Engineers and Scientists*. New York, USA: John Wiley and Sons Ltd., 1977.

[76] J. Makhoul, "Stable and efficient lattice methods for linear prediction," *IEEE Transactions on Acoustic Speech Signal Processing*, vol. 25, pp. 423–428, October 1977.

[77] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, pp. 561–580, April 1975.

[78] N. Jayant, "Adaptive quantization with a one-word memory," *Bell System Technical Journal*, vol. 52, pp. 1119–1144, September 1973.

[79] R. Steedman, "The common air interface MPT 1375," in Tuttlebee [626]. ISBN 3540196331.

[80] L. Hanzo, "The British cordless telephone system: CT2," in Gibson [622], ch. 29, pp. 462–477.

[81] H. Ochsner, "The digital european cordless telecommunications specification, DECT," in Tuttlebee [626], pp. 273–285. ISBN 3540196331.

[82] S. Asghar, "Digital European Cordless Telephone," in Gibson [622], ch. 30, pp. 478–499.

[83] "Personal handy phone (PHP) system." RCR Standard, STD-28, Japan.

[84] "CCITT recommendation G.721."

[85] N. Kitawaki, M. Honda, and K. Itoh, "Speech-quality assessment methods for speech coding systems," *IEEE Communications Magazine*, vol. 22, pp. 26–33, October 1984.

[86] A. Gray and J. Markel, "Distance measures for speech processing," *IEEE Transactions on Acoustic Speech Signal Processing*, vol. 24, no. 5, pp. 380–391, 1976.

[87] N. Kitawaki, H. Nagabucki, and K. Itoh, "Objective quality evaluation for low-bit-rate speech coding systems," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 242–249, February 1988.

[88] P. Noll and R. Zelinski, "Bounds on quantizer performance in the low bit-rate region," *IEEE Transactions on Communications*, pp. 300–304, February 1978.

[89] T. Thorpe, "The mean squared error criterion: Its effect on the performance of speech coders," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'89* [629], pp. 77–80.

[90] J. O'Neal, "Bounds on subjective performance measures for source encoding systems," *IEEE Transactions on Information Theory*, pp. 224–231, May 1971.

[91] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proceedings of the IEEE*, pp. 1551–1588, November 1985.

[92] B. Atal and M. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp. 247–254, June 1979.

[93] R. Steele, "Deploying personal communications networks," *IEEE Communications Magazine*, pp. 12–15, September 1990.

[94] J.-H. Chen, R. Cox, Y. Lin, N. Jayant, and M. Melchner, "A low-delay CELP codec for the CCITT 16 kb/s speech coding standard," *IEEE Journal on Selected Areas in Communications*, vol. 10, pp. 830–849, June 1992.

[95] D. Sen and W. Holmes, "PERCELP-perceptually enhanced random codebook excited linear prediction," in *Proceedings of IEEE Workshop on Speech Coding for Telecommunications*, pp. 101–102, 1993.

[96] S. Singhal and B. Atal, "Improving performance of multi-pulse LPC coders at low bit rates," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'84* [628], pp. 1.3.1–1.3.4.

[97] "Group speciale mobile (GSM) recommendation," April 1988.

[98] L. Hanzo and J. Stefanov, "The Pan-European Digital Cellular Mobile Radio System — known as GSM," in Steele and Hanzo [575], ch. 8, pp. 677–765.

[99] S. Singhal and B. Atal, "Amplitude optimization and pitch prediction in multipulse coders," *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp. 317–327, March 1989.

[100] "Federal standard 1016 – telecommunications: Analog to digital conversion of radio voice by 4,800 bits/second code excited linear prediction (CELP)," February 14 1991.

[101] S. Wang and A. Gersho, "Phonetic segmentation for low rate speech coding," in Atal *et al.* [52], pp. 257–266. ISBN: 0792390911.

[102] P. Lupini, H. Hassanein, and V. Cuperman, "A 2.4 kbit/s CELP speech codec with class-dependent structure," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'93)* [633], pp. 143–146.

[103] D. Griffin and J. Lim, "Multiband excitation vocoder," *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp. 1223–1235, August 1988.

[104] M. Nishiguchi, J. Matsumoto, R. Wakatsuki, and S. Ono, "Vector quantized MBE with simplified v/uv division at 3.0Kbps," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'93)* [633], pp. 151–154.

[105] W. Kleijn, "Encoding speech using prototype waveforms," *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 386–399, October 1993.

[106] V. Ramamoorthy and N. Jayant, "Enhancement of ADPCM speech by adaptive postfiltering," *Bell Systems Technical Journal*, vol. 63, pp. 1465–1475, October 1984.

[107] N. Jayant and V. Ramamoorthy, "Adaptive postfiltering of 16 kb/s-ADPCM speech," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'86*, (Tokyo, Japan), pp. 829–832, IEEE, 7–11 April 1986.

[108] J.-H. Chen and A. Gersho, "Real-time vector APC speech coding at 4800 bps with adaptive postfiltering," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'87* [631], pp. 2185–2188.

[109] ITU-T, *CCITT Recommendation G.728: Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction*, 1992.

[110] J.-H. Chen and A. Gersho, "Adaptive postfiltering for quality enhancement of coded speech," *IEEE Transactions on Speech and Audio Processing*, vol. 3, pp. 59–71, January 1995.

[111] F. Itakura and S. Saito, "Analysis-synthesis telephony based upon the maximum likelihood method," in *Proceedings of the 6th International Congress on Acoustic*, (Tokyo, Japan), pp. C17–20, 1968.

[112] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electronics and Communications in Japan*, vol. 53-A, pp. 36–43, 1970.

[113] N. Kitawaki, K. Itoh, and F. Itakura, "PARCOR speech analysis synthesis system," *Review of the Electronic Communication Lab., Nippon TTPC*, vol. 26, pp. 1439–1455, November-December 1978.

[114] R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE Transactions on Acoustic Speech Signal Processing*, pp. 309–321, 1975.

[115] N. Sugamura and N. Farvardin, "Quantizer design in LSP analysis-synthesis," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 432–440, February 1988.

[116] K. Paliwal and B. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 3–14, January 1993.

[117] F. Soong and B.-H. Juang, "Line spectrum pair (LSP) and speech data compression," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'84* [628], pp. 1.10.1–1.10.4.

[118] G. Kang and L. Fransen, "Low-bit rate speech encoders based on line-spectrum frequencies (LSFs)," Tech. Rep. 8857, NRL, November 1984.

[119] P. Kabal and R. Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials," *IEEE Transactions Acoustic Speech Signal Processing*, vol. 34, pp. 1419–1426, December 1986.

[120] M. Omologo, "The computation and some spectral considerations on line spectrum pairs (LSP)," in *Proceedings EUROSPEECH*, pp. 352–355, 1989.

[121] B. Cheetham, "Adaptive LSP filter," *Electronics Letters*, vol. 23, pp. 89–90, January 1987.

[122] K. Geher, *Linear Circuits*. Budapest, Hungary: Technical Publishers, 1972. (in Hungarian).

[123] N. Sugamura and F. Itakura, "Speech analysis and synthesis methods developed at ECL in NTT– from LPC to LSP," *Speech Communications*, vol. 5, pp. 199–215, June 1986.

[124] A. Lepschy, G. Mian, and U. Viaro, "A note on line spectral frequencies," *IEEE Transactions on Acoustic Speech Signal Processing*, vol. 36, pp. 1355–1357, August 1988.

[125] B. Cheetham and P. Huges, "Formant estimation from LSP coefficients," in *Proceedings IERE 5th International Conference on Digital Processing of Signals in Communications*, pp. 183–189, 20–23 September 1988.

[126] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Dordrecht: Kluwer Academic Publishers, 1992.

[127] Y. Shoham, "Vector predictive quantization of the spectral parameters for low rate speech coding," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'87* [631], pp. 2181–2184.

[128] R. Ramachandran, M. Sondhi, N. Seshadri, and B. Atal, "A two codebook format for robust quantisation of line spectral frequencies," *IEEE Transactions on Speech and Audio Processing*, vol. 3, pp. 157–168, May 1995.

[129] C. Xydeas and K. So, "Improving the performance of the long history scalar and vector quantisers," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'93)* [633], pp. 1–4.

[130] K. Lee, A. Kondoz, and B. Evans, "Speaker adaptive vector quantisation of LPC parameters of speech," *Electronic Letters*, vol. 24, pp. 1392–1393, October 1988.

[131] B. Atal, "Stochastic gaussian model for low-bit rate coding of LPC area parameters," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'87* [631], pp. 2404–2407.

[132] R. Salami, L. Hanzo, and D. Appleby, "A fully vector quantised self-excited vocoder," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'89* [629], pp. 124–128.

[133] M. Yong, G. Davidson, and A. Gersho, "Encoding of LPC spectral parameters using switched-adaptive interframe vector prediction," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'88* [632], pp. 402–405.

[134] J. Huang and P. Schultheis, "Block quantization of correlated gaussian random variables," *IEEE Transactions Communication Systems*, vol. 11, pp. 289–296, September 1963.

[135] R. Salami, L. Hanzo, and D. Appleby, "A computationally efficient CELP codec with stochastic vector quantization of LPC parameters," in *URSI International Symposium on Signals, Systems and Electronics*, (Erlangen, West Germany), pp. 140–143, 18–20 September 1989.

[136] B. Atal, R. Cox, and P. Kroon, "Spectral quantization and interpolation for CELP coders," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'89* [629], pp. 69–72.

[137] R. Laroia, N. Phamdo, and N. Farvardin, "Robust and efficient quantisation of speech LSP parameters using structured vector quantisers," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'91* [630], pp. 641–644.

[138] H. Harborg, J. Knudson, A. Fudseth, and F. Johansen, "A real time wideband CELP coder for a videophone application," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'94)* [615], pp. II121–II124.

[139] R. Lefebvre, R. Salami, C. Laflamme, and J. Adoul, "High quality coding of wideband audio signals using transform coded excitation (TCX)," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'94)* [615], pp. I193–I196.

[140] J. Paulus and J. Schnitzler, "16kbit/s wideband speech coding based on unequal subbands," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 255–258.

[141] J. Chen and D. Wang, "Transform predictive coding of wideband speech signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 275–278.

[142] A. Ubale and A. Gersho, "A multi-band CELP wideband speech coder," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)* [617], pp. 1367–1370.

[143] P. Combescure, J. Schnitzler, K. Fischer, R. Kirchherr, C. Lamblin, A. L. Guyader, D. Massaloux, C. Quinquis, J. Stegmann, and P. Vary, "A 16, 24, 32 Kbit/s wideband speech codec based on ATCELP," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'99*, IEEE, 1999.

[144] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *Journal of the Acoustic Society of America*, vol. 57, p. S35, 1975.

[145] L. Rabiner, M. Sondhi, and S. Levinson, "Note on the properties of a vector quantizer for LPC coefficients," *The Bell System Technical Journal*, vol. 62, pp. 2603–2616, October 1983.

[146] "7 khz audio coding within 64 kbit/s." CCITT Recommendation G.722, 1988.

[147] "Recommendation G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)." CCITT Study Group XVIII, 30 June 1995. Version 6.31.

[148] T. Eriksson, J. Linden, and J. Skoglung, "A safety-net approach for improved exploitation of speech correlation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'95)* [614], pp. 96–101.

[149] T. Eriksson, J. Linden, and J. Skoglung, "Exploiting interframe correlation in spectral quantization — a study of different memory VQ schemes," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 765–768.

[150] H. Zarrinkoub and P. Mermelstein, "Switched prediction and quantization of LSP frequencies," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 757–764.

[151] J. Natvig, "Evaluation of six medium bit-rate coders for the Pan-European digital mobile radio system," *IEEE Journal on Selected Areas in Communications*, pp. 324–331, February 1988.

[152] J. Schur, "Über potenzreihen, die im innern des einheitskreises beschränkt sind," *Journal für die reine und angewandte Mathematik, Bd 14*, pp. 205–232, 1917.

[153] W. Webb, L. Hanzo, R. Salami, and R. Steele, "Does 16-QAM provide an alternative to a half-rate GSM speech codec ?," in *Proceedings of IEEE Vehicular Technology Conference (VTC'91)*, (St. Louis, MO, USA), pp. 511–516, IEEE, 19–22 May 1991.

[154] L. Hanzo, W. Webb, R. Salami, and R. Steele, "On QAM speech transmission schemes for microcellular mobile PCNs," *European Transactions on Communications*, pp. 495–510, September/October 1993.

[155] J. Williams, L. Hanzo, R. Steele, and J. Cheung, "A comparative study of microcellular speech transmission schemes," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 909–925, November 1994.

[156] "Cellular system dual-mode mobile station-base station compatibility standard IS-54B." Telecommunications Industry Association Washington DC, 1992. EIA/TIA Interim Standard.

[157] Research and Development Centre for Radio Systems, Japan, *Public Digital Cellular (PDC) Standard, RCR STD-27*.

[158] R. Steele and L. Hanzo, eds., *Mobile Radio Communications*. New York, USA: IEEE Press-John Wiley, 2nd ed., 1999.

[159] L. Hanzo, W. Webb, and T. Keller, *Single- and Multi-carrier Quadrature Amplitude Modulation*. New York, USA: IEEE Press-John Wiley, April 2000.

[160] R. Salami, C. Laflamme, J.-P. Adoul, and D. Massaloux, "A toll quality 8 kb/s speech codec for the personal communications system (PCS)," *IEEE Transactions on Vehicular Technology*, pp. 808–816, August 1994.

[161] A. Black, A. Kondoz, and B. Evans, "High quality low delay wideband speech coding at 16 kbit/sec," in *Proceedings of 2nd International Workshop on Mobile Multimedia Communications*, 11–14 April 1995. Bristol University, UK.

[162] C. Laflamme, J.-P. Adoul, R. Salami, S. Morissette, and P. Mabilleau, "16 Kbps wideband speech coding technique based on algebraic CELP," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'91* [630], pp. 13–16.

[163] R. Salami, C. Laflamme, and J.-P. Adoul, "Real-time implementation of a 9.6 kbit/s ACELP wideband speech coder," in *Proceedings of GLOBECOM '92*, 1992.

[164] I. Gerson and M. Jasiuk, "Vector sum excited linear prediction (VSELP)," in Atal *et al.* [52], pp. 69–80. ISBN: 0792390911.

[165] M. Ireton and C. Xydeas, "On improving vector excitation coders through the use of spherical lattice codebooks (SLC's)," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'89* [629], pp. 57–60.

[166] C. Lamblin, J. Adoul, D. Massaloux, and S. Morissette, "Fast CELP coding based on the barnes-wall lattice in 16 dimensions," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'89* [629], pp. 61–64.

[167] C. Xydeas, M. Ireton, and D. Baghbadrani, "Theory and real time implementation of a CELP coder at 4.8 and 6.0 kbit/s using ternary code excitation," in *Proceedings of IERE 5th International Conference on Digital Processing of Signals in Communications*, pp. 167–174, September 1988.

[168] J. Adoul, P. Mabilleau, M. Delprat, and S. Morissette, "Fast CELP coding based on algebraic codes," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'87* [631], pp. 1957–1960.

[169] L. Hanzo and J. Woodard, "An intelligent multimode voice communications system for indoor communications," *IEEE Transactions on Vehicular Technology*, vol. 44, pp. 735–748, November 1995. ISSN 0018-9545.

[170] A. Kataoka, J.-P. Adoul, P. Combescure, and P. Kroon, "ITU-T 8-kbits/s standard speech codec for personal communication services," in *Proceedings of International Conference on Universal Personal Communications 1985*, (Tokyo, Japan), pp. 818–822, November 1995.

[171] H. Law and R. Seymour, "A reference distortion system using modulated noise," *IEE Paper*, pp. 484–485, November 1962.

[172] P. Kabal, J. Moncet, and C. Chu, "Synthesis filter optimization and coding: Applications to CELP," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'88* [632], pp. 147–150.

[173] Y. Tohkura, F. Itakura, and S. Hashimoto, "Spectral smoothing technique in PARCOR speech analysis-synthesis," *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp. 587–596, 1978.

[174] J.-H. Chen and R. Cox, "Convergence and numerical stability of backward-adaptive LPC predictor," in *Proceedings of IEEE Workshop on Speech Coding for Telecommunications*, pp. 83–84, 1993.

[175] S. Singhal and B. Atal, "Optimizing LPC filter parameters for multi-pulse excitation," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'83* [634], pp. 781–784.

[176] M. Fratti, G. Miani, and G. Riccardi, "On the effectiveness of parameter reoptimization in multipulse based coders," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'92* [635], pp. 73–76.

[177] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C*. Cambridge: Cambridge University Press, 1992.

[178] G. Golub and C. V. Loan, "An analysis of the total least squares problem," *SIAM Journal of Numerical Analysis*, vol. 17, no. 6, pp. 883–890, 1980.

[179] M. A. Rahham and K.-B. Yu, "Total least squares approach for frequency estimation using linear prediction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp. 1440–1454, 1987.

[180] R. Degroat and E. Dowling, "The data least squares problem and channel equalization," *IEEE Transactions on Signal Processing*, pp. 407–411, 1993.

[181] F. Tzeng, "Near-optimum linear predictive speech coding," in *IEEE Global Telecommunications Conference*, pp. 508.1.1–508.1.5, 1990.

[182] M. Niranjan, "CELP coding with adaptive output-error model identification," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'90* [623], pp. 225–228.

[183] J. Woodard and L. Hanzo, "Improvements to the analysis-by-synthesis loop in CELP codecs," in *Proceedings of IEE Conference on Radio Receivers and Associated Systems (RRAS'95)* [627], pp. 114–118.

[184] L. Hanzo, R. Salami, R. Steele, and P. Fortune, "Transmission of digitally encoded speech at 1.2 Kbaud for PCN," *IEE Proceedings, Part I*, vol. 139, pp. 437–447, August 1992.

[185] R. Cox, W. Kleijn, and P. Kroon, "Robust CELP coders for noisy backgrounds and noisy channels," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'89* [629], pp. 739–742.

[186] J. Campbell, V. Welch, and T. Tremain, "An expandable error-protected 4800 bps CELP coder (U.S. federal standard 4800 bps voice coder)," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'89* [629], pp. 735–738.

[187] S. Atungsiri, A. Kondoz, and B. Evans, "Error control for low-bit-rate speech communication systems," *IEE Proceedings-I*, vol. 140, pp. 97–104, April 1993.

[188] L. Ong, A. Kondoz, and B. Evans, "Enhanced channel coding using source criteria in speech coders," *IEE Proceedings-I*, vol. 141, pp. 191–196, June 1994.

[189] W. Kleijn, "Source-dependent channel coding and its application to CELP," in Atal *et al.* [52], pp. 257–266. ISBN: 0792390911.

[190] J. Woodard and L. Hanzo, "A dual-rate algebraic CELP-based speech transceiver," in *Proceedings of IEEE VTC '94* [619], pp. 1690–1694.

[191] C. Laflamme, J.-P. Adoul, H. Su, and S. Morissette, "On reducing the complexity of codebook search in CELP through the use of algebraic codes," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'90* [623], pp. 177–180.

[192] S. Nanda, D. Goodman, and U. Timor, "Performance of PRMA: A packet voice protocol for cellular systems," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 584–598, August 1991.

[193] M. Frullone, G. Riva, P. Grazioso, and C. Carciofy, "Investigation on dynamic channel allocation strategies suitable for PRMA schemes," *IEEE International Symposium on Circuits and Systems*, pp. 2216–2219, May 1993.

[194] J. Williams, L. Hanzo, and R. Steele, "Channel-adaptive voice communications," in *Proceedings of IEE Conference on Radio Receivers and Associated Systems (RRAS'95)* [627], pp. 144–147.

[195] W. Lee, "Estimate of channel capacity in Rayleigh fading environment," *IEEE Transactions on Vehicular Technology*, vol. 39, pp. 187–189, August 1990.

[196] T. Tremain, "The government standard linear predictive coding algorithm: LPC-10," *Speech Technology*, vol. 1, pp. 40–49, April 1982.

[197] J. Campbell, T. Tremain, and V. Welch, "The DoD 4.8 kbps standard (proprosed federal standard 1016)," in Atal *et al.* [52], pp. 121–133. ISBN: 0792390911.

[198] J. Marques, I. Trancoso, J. Tribolet, and L. Almeida, "Improved pitch prediction with fractional delays in CELP coding," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'90* [623], pp. 665–668.

[199] W. Kleijn, D. Kraisinsky, and R. Ketchum, "An efficient stochastically excited linear predictive coding algorithm for high quality low bit rate transmission of speech," *Speech Communication*, pp. 145–156, October 1988.

[200] Y. Shoham, "Constrained-stochastic excitation coding of speech at 4.8 kb/s," in Atal *et al.* [52], pp. 339–348. ISBN: 0792390911.

[201] A. Suen, J. Wand, and T. Yao, "Dynamic partial search scheme for stochastic codebook of FS1016 CELP coder," *IEE Proceedings*, vol. 142, no. 1, pp. 52–58, 1995.

[202] I. Gerson and M. Jasiuk, "Vector sum excited linear prediction (VSELP) speech coding at 8 kbps," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'90* [623], pp. 461–464.

[203] I. Gerson and M. Jasiuk, "Techniques for improving the performance of CELP-type speech codecs," *IEEEJournal on Selected Areas in Communications*, vol. 10, pp. 858–865, June 1992.

[204] I. Gerson, "Method and means of determining coefficients for linear predictive coding." US Patent No 544,919, October 1985.

[205] A. Cumain, "On a covariance-lattice algorithm for linear prediction," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'82* [636], pp. 651–654.

[206] W. Gardner, P. Jacobs, and C. Lee, "QCELP: a variable rate speech coder for CDMA digital cellular," in *Speech and Audio Coding for Wireless and Network Applications* (B. Atal, V. Cuperman, and A. Gersho, eds.), pp. 85–92, Dordrecht: Kluwer Academic Publishers, 1993.

[207] Telcomm. Industry Association (TIA), Washington, DC, USA, *Mobile station — Base station compatibility standard for dual-mode wideband spread spectrum cellular system, EIA/TIA Interim Standard IS-95*, 1993.

[208] T. Ohya, H. Suda, and T. Miki, "5.6 kbits/s PSI-CELP of the half-rate pdc speech coding standard," in *Proceedings of Vehicular Technology Conference*, vol. 3, (Stockholm), pp. 1680–1684, IEEE, May 1994.

[209] T. Ohya, T. Miki, and H. Suda, "Jdc half-rate speech coding standard and a real time operating prototype," *NTT Review*, vol. 6, pp. 61–67, November 1994.

[210] K. Mano, T. Moriya, S. Miki, H. Ohmuro, K. Ikeda, and J. Ikedo, "Design of a pitch synchronous innovation CELP coder for mobile communications," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 1, pp. 31–41, 1995.

[211] I. Gerson, M. Jasiuk, J.-M. Muller, J. Nowack, and E. Winter, "Speech and channel coding for the half-rate GSM channel," *Proceedings ITG-Fachbericht*, vol. 130, pp. 225–233, November 1994.

[212] A. Kataoka, T. Moriya, and S. Hayashi, "Implementation and performance of an 8-kbits/s conjugate structured CELP speech codec," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'94)* [615], pp. 93–96.

[213] R. Salami, C. Laflamme, and J.-P. Adoul, "8 kbits/s ACELP coding of speech with 10 ms speech frame: A candidate for CCITT standardization," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'94)* [615], pp. 97–100.

[214] J. Woodard, T. Keller, and L. Hanzo, "Turbo-coded orthogonal frequency division multiplex transmission of 8 kbps encoded speech," in *Proceeding of ACTS Mobile Communication Summit '97* [621], pp. 894–899.

[215] T. Ojanpare et al., "FRAMES multiple access technology," in *Proceedings of IEEE ISSSTA'96*, vol. 1, (Mainz, Germany), pp. 334–338, IEEE, September 1996.

[216] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo codes," in *Proceedings of the International Conference on Communications*, (Geneva, Switzerland), pp. 1064–1070, May 1993.

[217] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo codes," *IEEE Transactions on Communications*, vol. 44, pp. 1261–1271, October 1996.

[218] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on Information Theory*, vol. 42, pp. 429–445, March 1996.

[219] P. Jung and M. Nasshan, "Performance evaluation of turbo codes for short frame transmission systems," *IEE Electronic Letters*, pp. 111–112, January 1994.

[220] A. Barbulescu and S. Pietrobon, "Interleaver design for turbo codes," *IEE Electronic Letters*, pp. 2107–2108, December 1994.

[221] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimising symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

[222] "COST 207: Digital land mobile radio communications, final report." Office for Official Publications of the European Communities, 1989. Luxembourg.

[223] R. Salami, C. Laflamme, B. Bessette, and J.-P. Adoul, "Description of ITU-T recommendation G.729 annex A: Reduced complexity 8 kbits/s CS-ACELP codec," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)* [617], pp. 775–778.

[224] R. Salami, C. Laflamme, B. Bessette, and J.-P. Adoul, "ITU-T recommendation G.729 annex A: Reduced complexity 8 kbits/s CS-ACELP codec for digital simultaneous voice and data (DVSD)," *IEEE Communications Magazine*, vol. 35, pp. 56–63, September 1997.

[225] R. Salami, C. Laflamme, B. Besette, J.-P. Adoul, K. Jarvinen, J. Vainio, P. Kapanen, T. Hankanen, and P. Haavisto, "Description of the GSM enhanced full rate speech codec," in *Proceedings of ICC'97*, 1997.

[226] "PCS1900 enhanced full rate codec US1." SP-3612.

[227] "IS-136.1A TDMA cellular/PCS — radio interface — mobile station — base station compatibility digital control channel," August 1996. Revision A.

[228] T. Honkanen, J. Vainio, K. Jarvinen, P. Haavisto, R. Salami, C. Laflamme, and J. Adoul, "Enhanced full rate speech codec for IS-136 digital cellular system," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)* [617], pp. 731–734.

[229] "TIA/EIA/IS641, interim standard, TDMA cellular/PCS radio intergface — enhanced full-rate speech codec," May 1996.

[230] "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s." CCITT Recommendation G.723.1, March 1996.

[231] S. Bruhn, E. Ekudden, and K. Hellwig, "Adaptive Multi-Rate: A new speech service for GSM and beyond," in *Proceedings of 3rd ITG Conference on Source and Channel Coding*, (Technical Univ. Munich, Germany), pp. 319–324, 17th-19th, January 2000.

[232] A. Das, E. Paksoy, and A. Gersho, "Multimode and Variable Rate Coding of Speech," in *Speech Coding and Synthesis* (W. Kleijn and K. Paliwal, eds.), ch. 7, pp. 257–288, Elsevier, 1995.

[233] T. Taniguchi, S. Unagami, and R. Gray, "Multimode coding: a novel approach to narrow- and medium-band coding," *Journal of the Acoustic Society of America*, vol. 84, p. S12, 1988.

[234] P. Kroon and B. Atal, "Strategies for improving CELP coders," in *Proceedings of ICASSP*, vol. 1, pp. 151–154, 1988.

[235] A. D. Jaco, W. Gardner, P. Jacobs, and C. Lee, "QCELP: the North American CDMA digital cellular variable rate speech coding standard," in *Proceedings of IEEE Workshop on Speech Coding for Telecommunications*, pp. 5–6, 1993.

[236] E. Paksoy, K. Srinivasan, and A. Gersho, "Variable bit rate CELP coding of speech with phonetic classification," *European Transactions on Telecommunications*, vol. 5, pp. 591–602, Oct 1994.

[237] L. Cellario and D. Sereno, "CELP coding at variable rate," *European Transactions on Telecommunications*, vol. 5, pp. 603–613, Oct 1994.

[238] E. Yuen, P. Ho, and V. Cuperman, "Variable Rate Speech and Channel Coding for Mobile Communications," in *Proceedings of VTC, Stockholm, Sweden*, pp. 1709–1712, 8-10 June 1994.

[239] T. Kawashima, V. Sharma, and A. Gersho, "Capacity enhancement of cellular CDMA by traffic-based control of speech bit rate," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 543–550, August 1996.

[240] W. P. LeBlanc and S. A. Mahmoud, "Low Complexity, Low Delay Speech Coding for Indoor Wireless Communications," in *Proceedings of VTC, Stockholm, Sweden*, pp. 1695–1698, 8-10 June 1994.

[241] J. E. Kleider and W. M. Campbell, "An Adaptive Rate Digital Communication System For Speech," in *Proceedings of ICASSP, Munich, Germany*, vol. 3, pp. 1695–1698, 21-24th, April 1997.

[242] R. J. McAulay and T. F. Quatieri, "Low-rate Speech Coding Based on the Sinusoidal Model," in *Advances in Speech Signal Processing* (S. Furui and M. Sondhi, eds.), ch. 6, Marcel Dekker, New York, 1992.

[243] J. E. Kleider and R. J. Pattison, "Multi-Rate Speech Coding For Wireless and Internet Applications," in *Proceedings of ICASSP*, pp. 2379–2382, 15-19 March 1999.

[244] A. Das, A. D. Jaco, S. Manjunath, A. Ananthapadmanabhan, J. Juang, and E. Choy, "Multimode Variable Bit Rate Speech Coding: An Efficient Paradigm For High-Quality Low-Rate Representation Of Speech Signal," in *Proceedings of ICASSP*, 1999.

[245] "TIA/EIA/IS-127." Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems.

[246] F. Beritelli, A. Lombardo, S. Palazzo, and G. Schembra, "Performance Analysis of an ATM Multiplexer Loaded with VBR Traffic Generated by Multimode Speech Coders," *IEEE Journal On Selected Areas In Communications*, vol. 17, pp. 63–81, Jan 1999.

[247] "Description of a 12 kbps G.729 higher rate extension." ITU-T, Q19/16, Sept. 1997.

[248] L. Hanzo, W. Webb, and T. Keller, *Single- and Multi-carrier Quadrature Amplitude Modulation*. New York: John Wiley-IEEE Press, April 2000.

[249] S. Sampei, S. Komaki, and N. Morinaga, "Adaptive modulation/TDMA scheme for large capacity personal multi-media communication systems," *IEICE Transactions on Communications (Japan)*, vol. E77-B, pp. 1096–1103, September 1994.

[250] C. Wong and L. Hanzo, "Upper-bound performance of a wideband burst-by-burst adaptive modem," *IEEE Transactions on Communications*, vol. 48, pp. 367–369, March 2000.

[251] M. Yee, T. Liew, and L. Hanzo, "Radial basis function decision feedback equalisation assisted block turbo burst-by-burst adaptive modems," in *Proceeding of VTC'99 (Fall)* [611], pp. 1600–1604.

[252] E. Kuan and L. Hanzo, "Comparative study of adaptive-rate CDMA transmission employing joint-detection and interference cancellation receivers," in *accepted for the IEEE Vehicular Technology Conference, Tokyo, Japan*, 2000.

[253] E. L. Kuan, C. H. Wong, and L. Hanzo, "Burst-by-burst adaptive joint-detection CDMA," in *Proc. of IEEE VTC'99*, (Houston, USA), pp. 1628–1632, May 1999.

[254] T. Keller and L. Hanzo, "Sub–band adaptive pre–equalised OFDM transmission," in *Proceeding of VTC'99 (Fall)* [611], pp. 334–338.

[255] M. Münster, T. Keller, and L. Hanzo, "Co–channel interference suppression assisted adaptive OFDM in interference limited environments," in *Proceeding of VTC'99 (Fall)* [611], pp. 284–288.

[256] T. H. Liew, L. L. Yang, and L. Hanzo, "Soft-decision Redundant Residue Number System Based Error Correction Coding," in *Proc. of IEEE VTC'99*, (Amsterdam, The Netherlands), pp. 2546–2550, Sept 1999.

[257] S. Bruhn, P. Blocher, K. Hellwig, and J. Sjoberg, "Concepts and Solutions for Link Adaptation and Inband Signalling for the GSM AMR Speech Coding Standard," in *Proceedings of VTC, Houston, Texas, USA*, vol. 3, pp. 2451–2455, 16-19 May 1999.

[258] "GSM 05.09: Digital cellular telecommunications system (Phase 2+)." Link Adaptation, version 7.0.0, Release 1998.

[259] N. Szabo and R. Tanaka, *Residue Arithmetic and Its Applications to Computer Technology*. New York, USA: McGraw-Hill, 1967.

[260] F. Taylor, "Residue arithmetic: A tutorial with examples," *IEEE Computer Magazine*, vol. 17, pp. 50–62, May 1984.

[261] H. Krishna, K.-Y. Lin, and J.-D. Sun, "A coding theory approach to error control in redundant residue number systems - part I: Theory and single error correction," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 39, pp. 8–17, January 1992.

[262] J.-D. Sun and H. Krishna, "A coding theory approach to error control in redundant residue number systems - part II: Multiple error detection and correction," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 39, pp. 18–34, January 1992.

[263] L.-L. Yang and L. Hanzo, "Performance of residue number system based DS-CDMA over multipath fading channels using orthogonal sequences," *ETT*, vol. 9, pp. 525–536, November–December 1998.

[264] A. Klein, G. Kaleh, and P. Baier, "Zero forcing and minimum mean square error equalization for multiuser detection in code division multiple access channels," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 276–287, May 1996.

[265] W. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Transactions on Communications*, vol. 43, pp. 2223–2230, July 1995.

[266] A. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, pp. 1218–1230, October 1997.

[267] J. Torrance and L. Hanzo, "Upper bound performance of adaptive modulation in a slow Rayleigh fading channel," *Electronics Letters*, vol. 32, pp. 718–719, 11 April 1996.

[268] C. Wong and L. Hanzo, "Upper-bound of a wideband burst-by-burst adaptive modem," in *Proceeding of VTC'99 (Spring)* [613], pp. 1851–1855.

[269] M. Failli, "Digital land mobile radio communications COST 207," tech. rep., European Commission, 1989.

[270] C. Hong, *Low Delay Switched Hybrid Vector Excited Linear Predictive Coding of Speech*. PhD thesis, National University of Singapore, 1994.

[271] J. Zhang and H.-S. Wang, "A low delay speech coding system at 4.8 kb/s," in *Proceedings of the IEEE International Conference on Communications Systems*, vol. 3, pp. 880–883, November 1994.

[272] J.-H. Chen, N. Jayant, and R. Cox, "Improving the performance of the 16 kb/s LD-CELP speech coder," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'92* [635].

[273] J.-H. Chen and A. Gersho, "Gain-adaptive vector quantization with application to speech coding," *IEEE Transactions on Communications*, vol. 35, pp. 918–930, September 1987.

[274] J.-H. Chen and A. Gersho, "Gain-adaptive vector quantization for medium rate speech coding," in *Proceedings of IEEE International Conference on Communications 1985*, (Chicago, IL, USA), pp. 1456–1460, IEEE, 23–26 June 1985.

[275] J.-H. Chen, Y.-C. Lin, and R. Cox, "A fixed-point 16 kb/s LD-CELP algorithm," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'91* [630], pp. 21–24.

[276] J.-H. Chen, "High-quality 16 kb/s speech coding with a one-way delay less than 2 ms," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'90* [623], pp. 453–456.

[277] J. D. Marca and N. Jayant, "An algorithm for assigning binary indices to the codevectors of a multi-dimensional quantizer," in *Proceedings of IEEE International Conference on Communications 1987*, (Seattle, WA, USA), pp. 1128–1132, IEEE, 7–10 June 1987.

[278] K. Zeger and A. Gersho, "Zero-redundancy channel coding in vector quantization," *Electronic Letters*, vol. 23, pp. 654–656, June 1987.

[279] J. Woodard and L. Hanzo, "A low delay multimode speech terminal," in *Proceedings of IEEE VTC'96*, vol. 1, (Atlanta, GA, USA), pp. 213–217, IEEE, 28 April–1 May 1996.

[280] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantiser design," *IEEE Transactions on Communications*, vol. Com-28, January 1980.

[281] W. Kleijn, D. Krasinski, and R. Ketchum, "Fast methods for the CELP speech coding algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp. 1330–1342, August 1990.

[282] S. D'Agnoli, J. D. Marca, and A. Alcaim, "On the use of simulated annealing for error protection of CELP coders employing LSF vector quantizers," in *Proceedings of IEEE VTC '94* [619], pp. 1699–1703.

[283] X. Maitre, "7 kHz audio coding within 64 kbit/s," *IEEE Journal on Selected Areas of Communications*, vol. 6, pp. 283–298, February 1988.

[284] R. Crochiere, S. Webber, and J. Flanagan, "Digital coding of speech in sub-bands," *Bell System Technical Journal*, pp. 1069–1085, October 1976.

[285] R. Crochiere, "An analysis of 16 Kbit/s sub-band coder performance: dynamic range, tandem connections and channel errors," *Bell Systems Technical Journal*, vol. 57, pp. 2927–2952, October 1978.

[286] D. Esteban and C. Galand, "Application of quadrature mirror filters to split band voice coding scheme," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'77* [618], pp. 191–195.

[287] J. Johnston, "A filter family designed for use in quadrature mirror filter banks," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'80* [625], pp. 291–294.

[288] H. Nussbaumer, "Complex quadrature mirror filters," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'83* [634], pp. 221–223.

[289] C. Galand and H. Nussbaumer, "New quadrature mirror filter structures," *IEEE Transactions on Acoustic Speech Signal Processing*, vol. ASSP-32, pp. 522–531, June 1984.

[290] S. Quackenbush, "A 7 kHz bandwidth, 32 kbps speech coder for ISDN," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'91* [630], pp. 1–4.

[291] J. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas of Communications*, vol. 6, no. 2, pp. 314–323, 1988.

[292] E. Ordentlich and Y. Shoham, "Low-delay code-excited linear-predictive coding of wideband speech at 32kbps," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'91* [630], pp. 9–12.

[293] R. Soheili, A. Kondoz, and B. Evans, "New innovations in multi-pulse speech coding for bit rates below 8 kb/s," in *Proc. of Eurospeech*, pp. 298–301, 1989.

[294] V. Sanchez-Calle, C. Laflamme, R. Salami, and J.-P. Adoul, "Low-delay algebraic CELP coding of wideband speech," in *Signal Processing VI: Theories and Applications* (J. Vandewalle, R. Boite, M. Moonen, and A. Oosterlink, eds.), pp. 495–498, Netherlands: Elsevier Science Publishers, 1992.

[295] G. Roy and P. Kabal, "Wideband CELP speech coding at 16 kbit/sec," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'91* [630], pp. 17–20.

[296] R. Steele and W. Webb, "Variable rate QAM for data transmission over Rayleigh fading channels," in *Proceeedings of Wireless '91*, (Calgary, Alberta), pp. 1–14, IEEE, 1991.

[297] Y. Kamio, S. Sampei, H. Sasaoka, and N. Morinaga, "Performance of modulation-level-control adaptive-modulation under limited transmission delay time for land mobile communications," in *Proceedings of IEEE Vehicular Technology Conference (VTC'95)*, (Chicago, USA), pp. 221–225, IEEE, 15–28 July 1995.

[298] K. Arimochi, S. Sampei, and N. Morinaga, "Adaptive modulation system with discrete power control and predistortion-type non-linear compensation for high spectral efficient and high power efficient wireless communication systems," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'97* [624], pp. 472–477.

[299] M. Naijoh, S. Sampei, N. Morinaga, and Y. Kamio, "ARQ schemes with adaptive modulation/TDMA/TDD systems for wireless multimedia communication systems," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'97* [624], pp. 709–713.

[300] A. Goldsmith, "The capacity of downlink fading channels with variable rate and power," *IEEE Transactions on Vehicular Technolgy*, vol. 46, pp. 569–580, August. 1997.

[301] M.-S. Alouini and A. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *to appear IEEE Transactions on Vehicular Technology*, 1999. http://www.systems.caltech.edu.

[302] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Theory*, vol. 43, pp. 1986–1992, November 1997.

[303] T. Liew, C. Wong, and L. Hanzo, "Block turbo coded burst-by-burst adaptive modems," in *Proceedings of Microcoll'99, Budapest, Hungary*, pp. 59–62, 21–24 March 1999.

[304] C. Wong, T. Liew, and L. Hanzo, "Blind-detection assisted, block turbo coded, decision-feedback equalised burst-by-burst adaptive modulation." submitted to IEEE JSAC, 1999.

[305] H. Matsuoka, S. Sampei, N. Morinaga, and Y. Kamio, "Adaptive modulation system with variable coding rate concatenated code for high quality multi-media communications systems," in *Proceedings of IEEE VTC'96*, vol. 1, (Atlanta, GA, USA), pp. 487–491, IEEE, 28 April–1 May 1996.

[306] V. Lau and M. Macleod, "Variable rate adaptive trellis coded QAM for high bandwidth efficiency applications in rayleigh fading channels," in *Proceedings of IEEE Vehicular Technology Conference (VTC'98)* [620], pp. 348–352.

[307] A. Goldsmith and S. Chua, "Adaptive coded modulation for fading channels," *IEEE Transactions on Communications*, vol. 46, pp. 595–602, May 1998.

[308] T. Keller and L. Hanzo, "Adaptive orthogonal frequency division multiplexing schemes," in *Proceeding of ACTS Mobile Communication Summit '98* [612], pp. 794–799.

[309] E. Kuan, C. Wong, and L. Hanzo, "Burst-by-burst adaptive joint detection CDMA," in *Proceeding of VTC'99 (Spring)* [613].

[310] R. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *Bell Systems Technical Journal*, vol. 46, pp. 1775–1796, December 1966.

[311] L. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Transactions on Communications*, vol. 33, pp. 665–675, July 1985.

[312] K. Fazel and G. Fettweis, eds., *Multi-Carrier Spread-Spectrum*. Dordrecht: Kluwer, 1997. ISBN 0-7923-9973-0.

[313] T. May and H. Rohling, "Reduktion von Nachbarkanalstörungen in OFDM–Funkübertragungssystemen," in *2. OFDM–Fachgespräch in Braunschweig*, 1997.

[314] S. Müller and J. Huber, "Vergleich von OFDM–Verfahren mit reduzierter Spitzenleistung," in *2. OFDM–Fachgespräch in Braunschweig*, 1997.

[315] F. Classen and H. Meyr, "Synchronisation algorithms for an OFDM system for mobile communications," in *Codierung für Quelle, Kanal und Übertragung*, no. 130 in ITG Fachbericht, (Berlin), pp. 105–113, VDE–Verlag, 1994.

[316] F. Classen and H. Meyr, "Frequency synchronisation algorithms for OFDM systems suitable for communication over frequency selective fading channels," in *Proceedings of IEEE VTC '94* [619], pp. 1655–1659.

[317] S. Shepherd, P. van Eetvelt, C. Wyatt-Millington, and S. Barton, "Simple coding scheme to reduce peak factor in QPSK multicarrier modulation," *Electronics Letters*, vol. 31, pp. 1131–1132, July 1995.

[318] A. Jones, T. Wilkinson, and S. Barton, "Block coding scheme for reduction of peak to mean envelope power ratio of multicarrier transmission schemes," *Electronics Letters*, vol. 30, pp. 2098–2099, 1994.

[319] M. D. Benedetto and P. Mandarini, "An application of MMSE predistortion to OFDM systems," *IEEE Transactions on Communications*, vol. 44, pp. 1417–1420, November 1996.

[320] P. Chow, J. Cioffi, and J. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Transactions on Communications*, vol. 48, pp. 772–775, 1995.

[321] K. Fazel, S. Kaiser, P. Robertson, and M. Ruf, "A concept of digital terrestrial television broadcasting," *Wireless Personal Communications*, vol. 2, pp. 9–27, 1995.

[322] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Communications Magazine*, pp. 100–109, February 1995.

[323] J. Borowski, S. Zeisberg, J. Hübner, K. Koora, E. Bogenfeld, and B. Kull, "Performance of OFDM and comparable single carrier system in MEDIAN demonstrator 60GHz channel," in *Proceeding of ACTS Mobile Communication Summit '97* [621], pp. 653–658.

[324] I. Kalet, "The multitone channel," *IEEE Transactions on Communications*, vol. 37, pp. 119–124, February 1989.

[325] Y. Li and N. Sollenberger, "Interference suppression in OFDM systems using adaptive antenna arrays," in *Proceeding of Globecom'98*, (Sydney, Australia), pp. 213–218, IEEE, 8–12 November 1998.

[326] F. Vook and K. Baum, "Adaptive antennas for OFDM," in *Proceedings of IEEE Vehicular Technology Conference (VTC'98)* [620], pp. 608–610.

[327] T. Keller, J. Woodard, and L. Hanzo, "Turbo-coded parallel modem techniques for personal communications," in *Proceedings of IEEE VTC'97*, (Phoenix, AZ, USA), pp. 2158–2162, IEEE, 4–7 May 1997.

[328] T. Keller and L. Hanzo, "Blind-detection assisted sub-band adaptive turbo-coded OFDM schemes," in *Proceeding of VTC'99 (Spring)* [613], pp. 489–493.

[329] "Universal mobile telecommunications system (UMTS); UMTS terrestrial radio access (UTRA); concept evaluation," tech. rep., ETSI, 1997. TR 101 146.

[330] tech. rep. http://standards.pictel.com/ptelcont.htm#Audio or ftp://standard.pictel.com/sg16_q20/1999_09_Geneva/.

[331] M. Failli, "Digital land mobile radio communications COST 207," tech. rep., European Commission, 1989.

[332] J. Proakis, *Digital Communications*. New York, USA: McGraw-Hill, 3rd ed., 1995.

[333] H. Malvar, *Signal Processing with Lapped Transforms*. London, UK: Artech House, 1992.

[334] K. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages and Applications*. New York, USA: Academic Press Ltd., 1990.

[335] W. B. Kleijn and K. K. Paliwal, *Speech Coding and Synthesis*. Elsevier, 1995.

[336] P. Combescure, J. Schnitzler, K. Fischer, R. Kirchherr, C. Lamblin, A. L. Guyader, D. Massaloux, C. Quinquisd, J. Stegmann, and P. Vary, "A 16, 24, 32 Kbit/s Wideband Speech Codec Based on ATCELP," in *Proceedings of ICASSP, Phoenix, Arizona*, vol. 1, pp. 5–8, 14-18th, March 1999.

[337] J. Schnitzler, C. Erdmann, P. Vary, K. Fischer, J. Stegmann, C. Quinquis, D. Massaloux, and C. Lamblin, "Wideband Speech Coding For the GSM Adaptive Multi-Rate System," in *Proceedings of 3rd ITG Conference on Source and Channel Coding*, (Technical Univ. Munich, Germany), pp. 325–330, 17th-19th, January 2000.

[338] A. Murashima, M. Serizawa, and K. Ozawa, "A Multi-Rate Wideband Speech Codec Robust to Background Noise," in *Proceedings of ICASSP, Istanbul, Turkey*, vol. 2, pp. 1165–1168, 5-9th, June 2000.

[339] R. Steele and L. Hanzo, *(Editors), Mobile Radio Communications*. IEEE Press-John Wiley, 2nd Edition, 1999.

[340] R. V. Cox, J. Hagenauer, N. Seshadri, and C.-E. Sundberg, "Subband speech coding and matched convolutional channel coding for mobile radio channels," *IEEE Transactions on signal processing*, vol. 39, pp. 1717–1731, August 1991.

[341] J. Hagenauer, "Rate-compatible punctured convolutional codes (rcpc codes) and their applications," *IEEE Transactions on Communications*, vol. 36, pp. 389–400, April 1988.

[342] N. S. Othman, S. X. Ng, and L. Hanzo, "Turbo-dected unequal protection audio and speech transceivers using serially concatenated convolutional codes, trellis coded modulationa and space-time trellis coding," in *To appear in Proceedings of the IEEE Vehicular Technology Conference*, (Dallas, USA), 25-28 September 2005.

[343] S. X. Ng, J. Y. Chung, and L. Hanzo, "Turbo-dected unequal protection MPEG-4 telephony using trellis coded modulatioan and space-time trellis coding," in *Proceedings of IEE International Conference on 3G mobile communication Technologies (3G 2004)*, (London, UK), pp. 416–420, 18-20 October 2004.

[344] M. Tüchler and J. Hagenauer, "Exit charts of irregular codes," in *Proceedings of Conference on Information Science and Systems [CDROM]*, (Princeton University), 20-22 March 2002.

[345] M. Tüchler, "Design of serially concatenated systems depending on the block length," *IEEE Transactions on Communications*, vol. 52, pp. 209–218, February 2004.

[346] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Transactions on Communications*, vol. 49, pp. 1727–1737, October 2001.

[347] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec," *IEEE Transaction on Speech and Audio Processing*, vol. 10, pp. 620–636, November 2002.

[348] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimal symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

[349] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Transactions on Information Theory*, vol. 44, pp. 909–926, May 1998.

[350] A. Ashikhmin, G. Kramer, and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *IEEE Transactions on Information Theory*, vol. 50, pp. 2657– 2673, November 2004.

[351] I. Land, P. Hoeher, and S. Gligorević, "Computation of symbol-wise mutual information in transmission systems with logAPP decoders and application to EXIT charts," in *Proceedings of International ITG Conference on Source and Channel Coding (SCC)*, (Erlangen, Germany), pp. 195–202, January 2004.

[352] S. Dolinar and D. Divsalar, "Weight distributions for turbo codes using random and nonrandom permutations," *JPL-TDA Progress Report 42-122*, pp. 56–65, August 1995.

[353] A. Lillie, A. Nix, and J. McGeehan, "Performance and design of a reduced complexity iterative equalizer for precoded isi channels," in *Proceedings of IEEE Vehicular Technology Conference (VTC 2003-Fall)*, (Orlando, USA), 6-9 October 2003.

[354] "Information technology-coding of audio-visual objects-part3: Audio," *ISO/IEC 14496-3:2001*.

[355] "Specification for the use of video and audio coding in dvb services delivered directly over ip," *DVB Document A-84 Rev.1*, November 2005.

[356] "IP Datacast over DVB-H: Architecture," *DVB Document A098*, November 2005.

[357] "Extended AMR Wideband codec; Transcoding functions," *3GPP TS 26.290*.

[358] S. Ragot and B. Bessette and R. Lefebvre, "Low-complexity multi-rate lattice vector quantization with application to wideband speech coding at 32 kb/s," *Proceedings of ICASSP-2004*, May 2004.

[359] B. Bessette, R. Lefebvre, and R. Salami, "Universal speech/audio coding using hybrid acelp/tcx techniques," *Proc. ICASSP-2005*, March 2005.

[360] J. S. et al., "Rtp payload format for the extended adaptive multi-rate wideband (amr-wb+) audio codec," *IETF RFC 4352*, January 2006.

[361] "Method for the subjective assessment fo intermediate quality level of coding systems," *Recommendation ITU-R BS.1534*.

[362] "ETSI GSM 06.10." Full Rate (FR) speech transcoding (RPE-LTP: Regular Pulse Excitation - Long Term Prediction).

[363] "ETSI GSM 06.20." Half Rate (HR) speech transcoding (VSELP: Vector Sum Excited Linear Prediction.

[364] K. Jarvinen, J. Vainio, P. Kapanen, T. Honkanen, and P. Haavisto, "GSM Enhanced Full Rate Speech Codec," in *Proceedings of ICASSP, Munich, Germany*, vol. 2, pp. 771–774, 21-24th, April 1997.

[365] M. R. Schroeder and B. S. Atal, "Code Excited Linear Prediction (CELP) : High Quality Speech at Very Low Bit Rates," in *Proceedings of ICASSP*, (Tampa, Florida), pp. 937–940, March 1985.

[366] R. A. Salami, C. Laflamme, J. P. Adoul, and D. Massaloux, "A Toll Quality 8 kbit/s Speech Codec for the Personal Communications System(PCS)," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 808–816, Aug 1994.

[367] R. A. Salami and L. Hanzo, "Speech Coding," in *Mobile Radio Communications* (R. Steele and L. Hanzo, eds.), ch. 3, pp. 187–335, IEEE Press-John Wiley, 1999.

[368] L. Hanzo, F. C. A. Somerville, and J. Woodard, *Modern Speech Communications: Principles and Applications for Fixed and Wireless Channels*. John Wiley, 2001.

[369] F. Itakura, "Line Spectrum Representation of Linear Predictor Coefficients of Speech Signals," *Journal of Acoustical Society of America*, vol. 57, no. S35(A), 1975.

[370] K. K. Paliwal and B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 3–14, Jan 1993.

[371] L. Hanzo and J. P. Woodard, "An Intelligent Multimode Voice Communications System For Indoors Communications," *IEEE Transactions on Vehicular Technology*, vol. 44, pp. 735–749, Nov 1995.

[372] "Coding of Speech at 8 kbit/s using Conjugate-Structure Algebraic Code-Excited Linear Prediction (CS-ACELP)." ITU Recommendation G.729, 1995.

[373] J. H. Chen and A. Gersho, "Adaptive Postfiltering for Quality Enhancement of Coded Speech," *IEEE Transactions on Speech and Audio Processing*, vol. 3, pp. 59–71, Jan 1995.

[374] E. L. Kuan, C. H. Wong, and L. Hanzo, "Burst-by-burst adaptive joint-detection CDMA," in *Proc. of IEEE VTC'99*, vol. 2, (Houston, USA), pp. 1628–1632, May 1999.

[375] A. Das, E. Paksoy, and A. Gersho, "Multimode and Variable Rate Coding of Speech," in *Speech Coding and Synthesis* (W. Kleijn and K. Paliwal, eds.), ch. 7, pp. 257–288, Elsevier, 1995.

[376] T. Taniguchi, S. Unagami, and R. Gray, "Multimode coding: a novel approach to narrow- and medium-band coding," *Journal of the Acoustic Society of America*, vol. 84, p. S12, 1988.

[377] P. Kroon and B. Atal, "Strategies for improving CELP coders," in *Proceedings of ICASSP, New York, NY, USA*, vol. 1, pp. 151–154, 11-14th, April 1988.

[378] M. Yong and A. Gersho, "Vector excitation coding with dynamic bit allocation," in *Proceedings of IEEE Globecom, Hollywood, FL, USA*, pp. 290–294, Nov 28-Dec 1 1988.

[379] A. DeJaco, W. Gardner, P. Jacobs, and C. Lee, "QCELP: the North American CDMA digital cellular variable rate speech coding standard," in *Proceedings of IEEE Workshop on Speech Coding for Telecommunications*, pp. 5–6, 1993.

[380] E. Paksoy, K. Srinivasan, and A. Gersho, "Variable bit rate CELP coding of speech with phonetic classification," *European Transactions on Telecommunications*, vol. 5, pp. 591–602, Oct 1994.

[381] L. Cellario and D. Sereno, "CELP coding at variable rate," *European Transactions on Telecommunications*, vol. 5, pp. 603–613, Oct 1994.

[382] E. Yuen, P. Ho, and V. Cuperman, "Variable Rate Speech and Channel Coding for Mobile Communications," in *Proceedings of VTC, Stockholm, Sweden*, pp. 1709–1712, 8-10 June 1994.

[383] T. Kawashima, V. Sharma, and A. Gersho, "Capacity enhancement of cellular CDMA by traffic-based control of speech bit rate," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 543–550, Aug 1996.

[384] J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and their applications," *IEEE Transactions on Communications*, vol. 36, pp. 389–400, Apr 1988.

[385] W. P. LeBlanc and S. A. Mahmoud, "Low Complexity, Low Delay Speech Coding for Indoor Wireless Communications," in *Proceedings of VTC, Stockholm, Sweden*, pp. 1695–1698, 8-10 June 1994.

[386] "Coding of Speech at 16 kbit/s using Low-Delay Code Excited Linear Prediction (LD-CELP)." ITU Recommendation G.728, 1992.

[387] R. J. McAulay and T. F. Quatieri, "Low-rate Speech Coding Based on the Sinusoidal Model," in *Advances in Speech Signal Processing* (S. Furui and M. Sondhi, eds.), ch. 6, Marcel Dekker, New York, 1992.

[388] J. E. Kleider and R. J. Pattison, "Multi-Rate Speech Coding For Wireless and Internet Applications," in *Proceedings of ICASSP, Phoenix, Arizona*, pp. 1193–1196, 14-18th, March 1999.

[389] A. Das, A. Dejaco, S. Manjunath, A. Ananthapadmanabhan, J. Juang, and E. Choy, "Multimode Variable Bit Rate Speech Coding: An Efficient Paradigm For High-Quality Low-Rate Representation Of Speech Signal," in *Proceedings of ICASSP, Phoenix, Arizona*, pp. 3014–3017, 14-18th, March 1999.

[390] F. Beritelli, A. Lombardo, S. Palazzo, and G. Schembra, "Performance Analysis of an ATM Multiplexer Loaded with VBR Traffic Generated by Multimode Speech Coders," *IEEE Journal On Selected Areas In Communications*, vol. 17, pp. 63–81, Jan 1999.

[391] S. Sampei, S. Komaki, and N. Morinaga, "Adaptive Modulation/TDMA scheme for large capacity personal multimedia communications systems," *IEICE transactions on Communications*, vol. E77-B, pp. 1096–1103, September 1994.

[392] C. H. Wong and L. Hanzo, "Upper-bound performance of a wideband adaptive modem," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 367–369, 2000.

[393] M. S. Yee and L. Hanzo, "Block Turbo Coded Burst-By-Burst Adaptive Radial Basis Function Decision Feedback Equaliser Assisted Modems," in *Proc. of VTC'99 Fall, Amsterdam, Netherlands*, pp. 1600–1604, 19-22 September 1999.

[394] E. Kuan and L. Hanzo, "Comparative study of adaptive-rate CDMA transmission employing joint-detection and interference cancellation receivers," in *Proceedings of IEEE Vehicular Technology Conference, Tokyo, Japan*, vol. 1, pp. 71–75, 2000.

[395] T. Keller and L. Hanzo, "Sub-band Adaptive Pre-Equalised OFDM Schemes," in *Proceedings of the IEEE VTC'99 Fall, Amsterdam*, pp. 334–338, 1999.

[396] M. Muenster, T. Keller, and L. Hanzo, "Co-Channel Interference Suppression Assisted Adaptive OFDM in Interference-limited Environments," in *Proceedings of the IEEE VTC'99 Fall, Amsterdam*, pp. 284–288, 1999.

[397] T. H. Liew, L. L. Yang, and L. Hanzo, "Soft-decision Redundant Residue Number System Based Error Correction Coding," in *Proc. of IEEE VTC'99*, (Amsterdam, The Netherlands), pp. 2546–2550, Sept 1999.

[398] N. S. Szabo and R. I. Tanaka, *Residue Arithmetic and Its Applications to Computer Technology*. McGraw-Hill Book Company, 1967.

[399] F. J. Taylor, "Residue Arithmetic: A Tutorial with Examples," *IEEE Computer Magazine*, pp. 50–62, May 1984.

[400] H. Krishna, K. Y. Lin, and J. D. Sun, "A Coding Theory Approach to Error Control in Redundant Residue Number Systems - Part I: Theory and Single Error Correction," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 39, pp. 8–17, Jan 1992.

[401] J. D. Sun and H. Krishna, "A Coding Theory Approach to Error Control in Redundant Residue Number Systems - Part II: Multiple Error Detection and Correction," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 39, pp. 18–34, Jan 1992.

[402] L. L. Yang and L. Hanzo, "Performance of Residue Number System Based DS-CDMA over Multipath Channels Using Orthogonal Sequences," *European Transactions on Communications*, vol. 9, pp. 525–535, Nov.-Dec 1998.

[403] L. L. Yang and L. Hanzo, "Unified Error-Control Procedure for Global Telecommunication Systems Using Redundant Residue Number System Codes," in *Proc. of the WWRF (Wireless World Research Forum) Kick-off Meeting*, (Munich, Germany), 6-7th, March 2001.

[404] A. Klein, G. K. Kaleh, and P. W. Baier, "Zero forcing and minimum mean square error equalization for multiuser detection in code division multiple access channels," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 276–287, May 1996.

[405] W. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Transactions on Communications*, vol. 43, pp. 2223–2230, July 1995.

[406] A. J. Goldsmith and S. G. Chua, "Variable rate variable power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, pp. 1218–1230, October 1997.

[407] J. M. Torrance and L. Hanzo, "On the Upper bound performance of adaptive QAM in slow Rayleigh fading channel," *IEE Electronics Letters*, pp. 169–171, April 1996.

[408] C. H. Wong and L. Hanzo, "Upper-bound performance of a wideband burst-by-burst adaptive modem," in *Proc. of IEEE VTC'99*, (Houston, USA), pp. 1851–1855, May 1999.

[409] M. Faili, "Digital land mobile radio communications COST 207," tech. rep., European Commission, Luxembourg, 1989.

[410] L. Hanzo, F. C. A. Somerville, and J. P. Woodard, *Voice compression and communications: principles and applications for fixed and wireless channels*. Chichester, UK:John Wiley-IEEE Press, 2001.

[411] L. Hanzo, P. J. Cherriman, and J. Streit, *Wireless Video Communications: Second to Third Generation System and Beyond*. Piscataway, NJ:IEEE Press, 2001.

[412] L. Fielder, M. Bosi, G. Davidson, M. Davis, C. Todd, and S. Vernon, "AC-2 and AC-3: Low complexity transform-based audio coding," in *Collected Papers on Digital Audio Bit-Rate Reduction* (N. Gilchrist and C. Grewin, eds.), pp. 54–72, Audio Engineering Society, 1996.

[413] K. Tsutsui, H. Suzuki, O. Shimoyoshi, M. Sonohara, K. Akagiri, and R. Heddle, "ATRAC:adaptive transform acoustic coding for MiniDisc," in *Collected Papers on Digital Audio Bit-Rate Reduction* (N. Gilchrist and C. Grewin, eds.), pp. 95–101, Audio Engineering Society, 1996.

[414] J. Johnston, D. Sinha, S. Doward, and S. quackenbush, "AT&T perceptual audio coding (PAC)," in *Collected Papers on Digital Audio Bit-Rate Reduction* (N. Gilchrist and C. Grewin, eds.), pp. 73–81, Audio Engineering Society, 1996.

[415] G. C. P. Lokhoff, "Precision Adaptive Subband Coding (PASC) for the Digital Compact Cassette (DCC)," *IEEE Transactions on Consumer Electronic*, vol. 38, pp. 784–789, Nov 1992.

[416] P. Noll, "Digital audio coding for visual communications," *Proceedings of the IEEE*, vol. 83, pp. 925–943, June 1995.

[417] N. S. Jayant, J. Johnston, and R. Sofranek, "Signal Compression Based on Models of Human Perception," *Proceedings of IEEE*, vol. 81, pp. 1385–1422, Oct 1993.

[418] E. Zwicker, "Subdivision of the Audible Frequency Range Into Critical Bands (Frequenzgruppen)," *The Journal of the Acoustical Society of America*, vol. 33, p. 248, Feb 1961.

[419] K. Brandenburg, "Introduction to perceptual coding," in *Collected Papers on Digital Audio Bit-Rate Reduction* (N. Gilchrist and C. Grewin, eds.), pp. 23–30, Audio Engineering Society, 1996.

[420] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio," *Proceedings of the IEEE*, vol. 88, pp. 451–513, Apr 2000.

[421] H. Fletcher, "Auditory patterns," *Rev. Mod. Phys.*, pp. 47–65, Jan 1940.

[422] D. D. Greenwood, "Critical bandwidth and the frequency coordinates of the Basilar membrane," *The Journal of the Acoustical Society of America*, pp. 1344–1356, Oct 1961.

[423] B. Scharf, "Critical bands," in *Foundations of Modern Auditory Theory*, New York, Academic Publisher, 1970.

[424] E. Zwicker and H. Fastl, *Psychoacoustics - Facts and Models*. Springer-Verlag, Berlin, Germany, 1990.

[425] R. Hellman, "Asymmetry of masking between noise and tone," *Percep. Psychphys.*, vol. 11, pp. 241–246, 1972.

[426] K. Brandenburg, "OCF - A new coding algorithm for high quality sound signals," in *Proceedings of ICASSP, Dallas, Texas, USA*, pp. 141–144, April 1987.

[427] J. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 314–323, Feb 1988.

[428] D. Esteban and C. Galand, "Application of Quadrature Mirror Filters to Split Band Voice Coding Scheme," in *Proceedings of ICASSP*, pp. 191–195, 9-11th, May 1977.

[429] R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital Coding of Speech in Subbands," *The Bell System Technical Journal*, vol. 55, pp. 1069–1085, Oct 1976.

[430] J. M. Tribolet and R. E. Crochiere, "Frequency Domain Coding of Speech," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, pp. 512–530, Oct 1979.

[431] N. S. Jayant and P. Noll, *Digital Coding of Waveforms : Principles and Applications to Speech and Video Coding*. Prentice Hall, INC. Englewood Cliffs, NJ, 1984.

[432] R. Zelinski and P. Noll, "Adaptive Transform Coding of Speech Signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, pp. 299–309, Aug 1977.

[433] J. Herre and J. Johnston, "Continuously signal-adaptive filterbank for high-quality perceptual audio coding," in *Proceedings of IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 19-22nd, Oct 1997.

[434] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks and applications: a tutorial," *Proceedings of the IEEE*, vol. 78, pp. 56–93, Jan 1990.

[435] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[436] A. Akansu and M. T. J. S. (Eds), *Subband and Wavelet Transforms, Designs and Applications*. Norwell, MA:Kluwer Academic, 1996.

[437] H. S. Malvar, *Signal Processing with Lapped Transforms*. Boston, MA: Artech House, 1992.

[438] H. S. Malvar, "Modulated QMF filter banks with perfect reconstruction," *Electronics Letters*, vol. 26, pp. 906–907, June 1990.

[439] J. H. Rothweiler, "Polyphase quadrature filters: A new subband coding technique," in *Proceedings of ICASSP, Boston, Massachusetts, USA*, pp. 1280–1283, April 1983.

[440] M. Temerinac and B. Edler, "LINC: A Common Theory of Transform and Subband Coding," *IEEE Transactions on Communications*, vol. 41, no. 2, pp. 266–274, 1993.

[441] H. J. Nussbaumer, "Pseudo QMF filter bank," *IBM Tech. Disclosure Bull.*, vol. 24, pp. 3081–3087, Nov 1981.

[442] J. P. Princen and A. B. Bradley, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-34, pp. 1153–1161, Oct 1986.

[443] S. Shlien, "The Modulated Lapped Transfrom, Its Time-Varying Forms, and Its Applications to Audio Coding Standards," *IEEE Transactions on Speeech and Audio Processing*, vol. 5, pp. 359–366, July 1997.

[444] B. Edler, "Coding of Audio Signals with Overlapping Block Transform and Adaptive Window Functions," *(in German)*, *Frequenz*, vol. 43, pp. 252–256, Sept 1989.

[445] K. Brandenburg and G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio," in *Collected Papers on Digital Audio Bit-Rate Reduction* (N. Gilchrist and C. Grewin, eds.), pp. 31–42, Audio Engineering Society, 1996.

[446] H. S. Malvar, "Lapped Transform for Efficient Transform/Subband Coding," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 969–978, June 1990.

[447] Y. Mahieux, J. Petit, and A. Charbonnier, "Transform coding of audio signals using correlation between succesive transform blocks," in *Proceedings of ICASSP, Glasgow, Scotland*, pp. 2021–2024, May 1989.

[448] J. Johnston and A. Ferreira, "Sum-difference stereo transform coding," in *Proceedings of ICASSP, San Fransisco, CA, USA*, pp. II–569–II–572, March 1992.

[449] S. Park, Y. Kim, and Y. Seo, "Multi-Layer Bit-Sliced Bit-Rate Scalable Audio Coding," in *103th Convention of the Audio Engineering Society*, preprint 4520, Sep 1997.

[450] N. Iwakami, T. Moriya, and S. Miki, "High-quality audio coding at less than 64 kbps by using transform-domain weighted interleave vector quantization (TwinVQ)," in *Proceedings of ICASSP, Detroit, MI, USA*, pp. 3095–3098, 9-12th, May 1995.

[451] J. Herre and J. Johnston, "Enhancing the performance of perceptual audio coders by using temporal noise shaping (TNS)," in *101th Convention of the Audio Engineering Society*, preprint 4384, Dec 1996.

[452] J. Herre, K. Brandenburg, and D. Lederer, "Intensity Stereo Coding," in *96th Convention of the Audio Engineering Society*, preprint 3799, May 1994.

[453] J. Johnston, "Perceptual transform coding of wideband stereo signals," in *Proceedings of ICASSP, Glasgow, Scotland*, pp. 1993–1996, May 1989.

[454] R. G. van der Waal and R. N. J. Veldhuis, "Subband coding of stereophonic digital audio signals," in *Proceedings of ICASSP, Toronto, Canada*, vol. 5, pp. 3601–3604, 14-17th, May 1991.

[455] T. V. Sreenivas and M. Dietz, "Vector Quantization of Scale Factors in Advanced Audio Coder (AAC)," in *Proceedings of ICASSP, Seattle, Washington, USA*, vol. 6, pp. 3641–3644, 12-15th, May 1998.

[456] D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," *Proceedings of IRE*, vol. 40, pp. 1098–1101, Sept 1952.

[457] ISO/IEC JTC1/SC29/WG11/N2203TF, MPEG-4 Audio Version 1 Final Committee Draft 14496-3 Subpart 4:TF, `http://www.tnt.uni-hannover.de/project/mpeg/audio/documents/`, May 1998.

[458] K. Ikeda, T. Moriya, N. Iwakami, A. Jin, and S. Miki, "A design of TwinVQ audio codec for personal communication systems," in *Fourth IEEE International Conference on Universal Personal Communications.*, pp. 803–807, 1995.

[459] K. Ikeda, T. Moriya, and N. Iwakami, "Error protected TwinVQ audio coding at less than 64 kbit/s," in *Proceedings of IEEE Speech Coding Workshop*, pp. 33–34, 1995.

[460] T. Moriya, N. Iwakami, K. Ikeda, and S. Miki, "Extension and complexity reduction of TwinVQ audio coder," in *Proceedings of ICASSP, Atlanta, Georgia*, pp. 1029–1032, 7-10th, May 1996.

[461] T. Moriya, "Two-channel conjugate vector quantizer for noisy channel speech coding," *IEEE Journal on Selected Areas in Communications*, vol. 10, pp. 866–874, 1992.

[462] N. Kitawaki, T. Moriya, T. Kaneko, and N. Iwakami, "Comparison of two speech and audio coders at 8 kbit/s from the viewpoints of coding scheme and quality," *IEICE Transactions on Communications*, vol. E81-B, pp. 2007–2012, Nov 1998.

[463] T. Moriya, N. Iwakami, A. Jin, K. Ikeda, and S. Miki, "A design of transform coder for both speech and audio signals at 1 bit/sample," in *Proceedings of ICASSP, Munich, Germany*, pp. 1371–1374, 21-24th, April 1997.

[464] T. Moriya and M. Honda, "Transform coding of speech using a weighted vector quantizer," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 425–431, Feb 1988.

[465] H. Purnhagen, "An Overview of MPEG-4 Audio Version 2," in *AES 17th International Conference on High-Quality Audio Coding*, Sept 1999.

[466] H. Purnhagen, "Advances in parametric audio coding," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York*, pp. W99–1–W99–4, 17-20th, Oct 1999.

[467] H. Purnhagen and N. Meine, "HILN - The MPEG-4 Parametric Audio Coding Tools," in *Proceedings - IEEE International Symposium on Circuits and Systems, Geneva, Switzerland*, pp. III–201–III–204, 29-31th, May 2000.

[468] B. Edler and H. Purnhagen, "Concepts for Hybrid Audio Coding Schemes Based on Parametric Techniques," in *AES 105th Convention*, Preprint 4808, Sept 1998.

[469] S. Levine, T. Verma, and J. O. Smith, "Multiresolution sinusoidal modelling for wideband audio with modifications," in *Proceedings of ICASSP, Seattle, Washington, USA*, vol. 6, pp. 3585–3588, 12-15th, May 1998.

[470] T. S. Verma and T. H. Y. Meng, "Analysis/synthesis tool for transient signals that allows a flexible sines+transients+noise model for audio," in *Proceedings of ICASSP, Seattle, Washington, USA*, vol. 6, pp. 3573–3576, 12-15th, May 1998.

[471] S. Levine and J. O. S. III, "A switched parametric & transform audio coder," in *Proceedings of ICASSP, Phoenix, Arizona*, pp. 985–988, 14-18th, March 1999.

[472] M. Nishiguchi and J. Matsumoto, "Harmonic and Noise Coding of LPC Residuals with Classified Vector Quantization," in *Proceedings of ICASSP, Detroit, MI, USA*, vol. 1, pp. 484–487, 9-12th, May 1995.

[473] M. Nishiguchi, K. Iijima, and J. Matsumoto, "Harmonic vector excitation coding of speech at 2 kbit/s," in *Proceedings of IEEE Workshop on Speech Coding for Telecommunications*, pp. 39–40, 1997.

[474] ISO/IEC JTC1/SC29/WG11/N2203PAR, MPEG-4 Audio Version 1 Final Committee Draft 14496-3 Subpart 2:Parametric Coding, `http://www.tnt.uni-hannover.de/project/mpeg/audio/documents/`, March 1998.

[475] ISO/IEC JTC1/SC29/WG11/N2203CELP, MPEG-4 Audio Version 1 Final Committee Draft 14496-3 Subpart 3:CELP, `http://www.tnt.uni-hannover.de/project/mpeg/audio/documents/`, May 1998.

[476] R. Taori, R. J. Sluijter, and A. J. Gerrits, "On scalability in CELP coding systems," in *Proceedings of IEEE Workshop on Speech Coding for Telecommunications*, pp. 67–68, Sep 1997.

[477] K. Ozawa, M. Serizawa, and T. Nomura, "High quality MP-CELP speech coding at 12 kb/s and 6.4 kb/s," in *Proceedings of IEEE Workshop on Speech Coding for Telecommunications*, pp. 71–72, Sep 1997.

[478] E. F. Deprettere and P. Kroon, "Regular excitation reduction for effective and efficient LP-coding of speech," in *Proceedings of ICASSP*, (Tampa, Florida), pp. 965–968, March 1985.

[479] C. Laflamme, J. P. Adoul, R. A. Salami, S. Morissette, and P. Mabilleau, "16kbit/s Wideband Speech Coding Technique Based On Algebraic CELP," in *Proceedings of ICASSP, Toronto, Canada*, vol. 1, pp. 13–16, 14-17th, May 1991.

[480] P. Kroon, R. J. Sluyter, and E. F. Deprettere, "Regular-pulse excitation - a novel approach to effective and efficient multipulse coding of speech," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, pp. 1054–1063, Oct 1986.

[481] P. Chaudhury, "The 3GPP proposal for IMT-2000," *IEEE Communications Magazine*, vol. 37, pp. 72–81, Dec 1999.

[482] G. Foschini Jr. and M. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, March 1998.

[483] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.

[484] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, pp. 1456–1467, July 1999.

[485] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 451–460, March 1999.

[486] G. Bauch, "Concatenation of space-time block codes and Turbo-TCM," in *Proceedings of IEEE International Conference on Communications, Vancouver, Canada*, pp. 1202–1206, June 1999.

[487] D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri, "Space-time coded OFDM for high data-rate wireless communication over wideband channels," in *Proceedings of IEEE Vehicular Technology Conference*, (Ottawa, Canada), pp. 2232–2236, May 1998.

[488] Y. Li, J. Chuang, and N. Sollenberger, "Transmitter diversity for OFDM systems and its impact on high-rate data wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1233–1243, July 1999.

[489] A. Naguib, N. Seshadri, and A. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Magazine*, vol. 17, pp. 76–92, May 2000.

[490] A. Naguib, V. Tarokh, N. Seshadri, and A. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1459–1478, October 1998.

[491] H. Holma and A. Toskala, *WCDMA for UMTS*. John Wiley-IEEE Press, April 2000.

[492] R. W. Chang, "Synthesis of Band-Limited Orthogonal Signals for Multichannel Data Transmission," *BSTJ*, vol. 46, pp. 1775–1796, Dec 1966.

[493] J. Cimini, "Analysis and Simulation of a Digital Mobile Channel Using Orthogonal Frequency Division Multiplexing," *IEEE Transactions on Communications*, vol. COM-33, pp. 665–675, July 1985.

[494] K. Fazel and G. Fettweis, *Multi-carrier spread spectrum*. Kluwer, 1997.

[495] P. R. K. Fazel, S. Kaiser, and M. Ruf, "A concept of digital terrestial television broadcasting," *Wireless Personal Communications*, vol. 2, pp. 9–27, 1995.

[496] I. J. H. Sari and G. Karam, "Transmission techniques for digital terrestial tv broadcasting," *IEEE Communications Magazine*, vol. 33, pp. 100–109, Feb 1995.

[497] J. Borowski, S. Zeiberg, J. Hbner, K. Koora, E. Bogenfeld, and B. Kull, "Performance of ofdm and comparable single carrier system in median demonstration 60 ghz channel," in *Proceedings of ACTS Summit*, (Aalborg, Denmark), pp. 653–658, Oct 1997.

[498] J. C. I. Chuang, Y. G. Li, and N. R. Sollenberger, "OFDM based High-speed Wireless Access for Internet Applications," in *Proceedings of PIMRC Fall, London, UK*, 18-21th September 2000.

[499] T. Keller, M. Muenster, and L. Hanzo, "A Turbo-Coded Burst-By-Burst Adaptive Wideband Speech Transceiver," *Journal on Selected Areas in Communications*, vol. 18, pp. 2363–2372, Nov 2000.

[500] L. Hanzo, C. H. Wong, and P. Cherriman, "Channel-adaptive wideband wireless video telephony," *IEEE Signal Processing Magazine*, vol. 17, pp. 10–30, July 2000.

[501] MPEG Audio Web Page, `http://www.tnt.uni-hannover.de/project/mpeg/audio/`.

[502] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo codes," in *IEEE International Conference on Communications*, pp. 1064–1070, 23-26th, May 1993.

[503] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

[504] R. Koenen , "MPEG-4 Overview," in *ISO/IEC JTC1/SC29/WG11 N4668, version 21-Jeju Version, ISO/IEC*, (http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm), March 2002.

[505] R. Koenen, "MPEG-4 Multimedia for Our Time," vol. 36, pp. 26–33, February 1999.

[506] ISO/IEC JTC1/SC29/WG11 N2503, "Information Technology-Very Low Bitrate Audio-Visual Coding," in *ISO/IEC 14496-3. Final Draft International Standard. Part 3: Audio*, 1998.

[507] J. Herre, and B. Grill, "Overview of MPEG-4 Audio and its Applications in Mobile Communications," vol. 1, pp. 11–20, August 2000.

[508] F. Pereira and T. Ebrahimi , *The MPEG-4 Book*. New Jersey, USA: Prentice Hall PTR IMSC Press, 2002.

[509] S. Lin and D. J. Costello, Jr, *Error Control Coding: Fundamentals and Applications*. Inc. Englewood Cliffs, New Jersey 07632: Prentice-Hall, 1983.

[510] G. Ungerböck, "Channel Coding with Multilevel/Phase Signals," *IEEE Transactions on Information Theory*, vol. 28, pp. 55–67, January 1982.

[511] L. Hanzo, T. H. Liew and B. L. Yeap, *Turbo Coding, Turbo Equalisation and Space Time Coding for Transmission over Wireless channels*. New York, USA: John Wiley IEEE Press, 2002.

[512] L. Hanzo, S. X. Ng, W. T. Webb and T. Keller, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. New York, USA: John Wiley IEEE Press, 2004.

[513] V. Tarokh, N. Seshadri and A. R. Calderbank, "Space-time Codes for High Rate Wireless Communication: Performance analysis and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.

[514] L. Hanzo, P.J. Cherriman and J. Street, *Wireless Video Communications: Second to Third Generation Systems and Beyond*. NJ, USA : IEEE Press., 2001.

[515] L. Hanzo, F.C.A. Somerville, and J.P. Woodard, *Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels*. Chichester, UK: John Wiley-IEEE Press, 2001.

[516] S. X. Ng, J. Y. Chung and L. Hanzo, "Turbo-Detected Unequal Protection MPEG-4 Wireless Video Telephony using Trellis Coded Modulation and Space-Time Trellis Coding," in *IEE International Conference on 3G Mobile Communication Technologies (3G 2004)*, (London, UK), 18 - 20 October 2004.

[517] E. Zehavi, "8-PSK trellis codes for a Rayleigh fading channel," *IEEE Transactions on Communications*, vol. 40, pp. 873–883, May 1992.

[518] S. X. Ng, J. Y. Chung and L. Hanzo, "Integrated wireless multimedia turbo-transceiver design - Interpreting Shannon's lessons in the turbo-era," in *IEE Sparse-Graph Codes Seminar*, (The IEE, Savoy Place, London), October 2004.

[519] R.V. Cox, J. Hagenauer, N. Seshadri, and C-E. W. Sundberg, "Subband Speech Coding and Matched Convolutional Coding for Mobile Radio Channels," *IEEE Transaction on Signal Processing*, vol. 39, pp. 1717–1731, August 1991.

[520] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola and K. Jarvinen, "The Adaptive Multirate Wideband Speech Codec (AMR-WB)," *IEEE Transactions on Speech and Audio Processing*, vol. 10, pp. 620–636, November 2002.

[521] 3GPP TS 26.173, "Adaptive Multi-Rate Wideband Speech ANSI-C Code," in *3GPP Technical Specification*, 2003.

[522] S. X. Ng and L. Hanzo, "On the MIMO Channel Capacity of Multi-Dimensional Signal Sets," in *IEEE Vehicular Technology Conference*, (Los Angeles, USA), 26-29 September 2004.

[523] T. Fingscheidt, and P. Vary, "Softbit Speech Decoding: A New Approach to Error Concealment," *IEEE Transaction on Speech and Audio Processing*, vol. 9, pp. 240–251, March 2001.

[524] B. Atal and M. Schroeder, "Predictive coding of speech signals," *Bell System Technical Journal*, pp. 1973–1986, October 1970.

[525] I. Wassel, D. Goodman, and R. Steele, "Embedded delta modulation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, pp. 1236–1243, August 1988.

[526] B. Atal and S. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *The Journal of the Acoustical Society of America*, vol. 50, no. 2, pp. 637–655, 1971.

[527] M. Kohler, L. Supplee, and T. Tremain, "Progress towards a new government standard 2400bps voice coder," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'95)* [614], pp. 488– –491.

[528] K. Teague, B. Leach, and W. Andrews, "Development of a high-quality MBE based vocoder for implementation at 2400bps," in *Proceedings of the IEEE Wichita Conference on Communications, Networking and Signal Processing*, pp. 129–133, April 1994.

[529] H. Hassanein, A. Brind'Amour, S. Déry, and K. Bryden, "Frequency selective harmonic coding at 2400bps," in *Proceedings of the 37th Midwest Symposium on Circuits and Systems*, vol. 2, pp. 1436–1439, 1995.

[530] R. McAulay and T. Quatieri, "The application of subband coding to improve quality and robustness of the sinusoidal transform coder," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'93)* [633], pp. 439–442.

[531] A. McCree and T. Barnwell III, "A mixed excitation LPC vocoder model for low bit rate speech coding," *IEEE Transactions on Speech and audio Processing*, vol. 3, no. 4, pp. 242–250, 1995.

[532] P. Laurent and P. L. Noue, "A robust 2400bps subband LPC vocoder," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'95)* [614], pp. 500–503.

[533] W. Kleijn and J. Haagen, "A speech coder based on decomposition of characteristic waveforms," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'95)* [614], pp. 508–511.

[534] R. McAulay and T. Champion, "Improved interoperable 2.4 kb/s LPC using sinusoidal transform coder techniques," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'90* [623], pp. 641–643.

[535] K. Teague, W. Andrews, and B. Walls, "Harmonic speech coding at 2400 bps," in *Proceedings of 10th Annual Mid-America Symposium on Emerging Computer Technology*, (Norman, Oklahoma, USA), 1996.

[536] J. Makhoul, R. Viswanathan, R. Schwartz, and A. Huggins, "A mixed-source model for speech compression and synthesis," *The Journal of the Acoustical Society of America*, vol. 64, no. 4, pp. 1577–1581, 1978.

[537] A. McCree, K. Truong, E. George, T. Barnwell, and V. Viswanathan, "A 2.4kbit/s coder candidate for the new U.S. federal standard," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 200–203.

[538] A. McCree and T. Barnwell III, "Improving the performance of a mixed excitation LPC vocoder in acoustic noise," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'92* [635], pp. 137–140.

[539] J. Holmes, "The influence of glottal waveform on the naturalness of speech from a parallel formant synthesizer," *IEEE Transaction on Audio and Electroacoustics*, vol. 21, pp. 298–305, June 1973.

[540] W. Kleijn, Y. Shoham, D. Sen, and R. Hagen, "A low-complexity waveform interpolation coder," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 212–215.

[541] D. Hiotakakos and C. Xydeas, "Low bit rate coding using an interpolated zinc excitation model," in *Proceedings of the ICCS 94*, pp. 865–869, 1994.

[542] R. Sukkar, J. LoCicero, and J. Picone, "Decomposition of the LPC excitation using the zinc basis functions," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 9, pp. 1329–1341, 1989.

[543] M. Schroeder, B. Atal, and J. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *Journal of the Acoustical Society of America*, vol. 66, pp. 1647–1652, December 1979.

[544] W. Voiers, "Diagnostic acceptability measure for speech communication systems," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'77* [618], pp. 204–207.

[545] W. Voiers, "Evaluating processed speech using the diagnostic rhyme test," *Speech Technology*, January/February 1983.

[546] T. Tremain, M. Kohler, and T. Champion, "Philosophy and goals of the D.O.D 2400bps vocoder selection process," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 1137–1140.

[547] M. Bielefeld and L. Supplee, "Developing a test program for the DoD 2400bps vocoder selection process," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 1141–1144.

[548] J. Tardelli and E. Kreamer, "Vocoder intelligibility and quality test methods," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 1145–1148.

[549] A. Schmidt-Nielsen and D. Brock, "Speaker recognizability testing for voice coders," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 1149–1152.

[550] E. Kreamer and J. Tardelli, "Communicability testing for voice coders," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 1153–1156.

[551] B. Atal and L. Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition," *IEEE Transactions an Acoustics, Speech and Signal Processing*, vol. 24, pp. 201–212, June 1976.

[552] T. Ghiselli-Crippa and A. El-Jaroudi, "A fast neural net training algorithm and its application to speech classification," *Engineering Applications of Artificial Intelligence*, vol. 6, no. 6, pp. 549–557, 1993.

[553] A. Noll, "Cepstrum pitch determination," *Journal of the Acoustical Society of America*, vol. 41, pp. 293–309, February 1967.

[554] S. Kadambe and G. Boudreaux-Bartels, "Application of the wavelet transform for pitch detection of speech signals," *IEEE Transactions on Information Theory*, vol. 38, pp. 917–924, March 1992.

[555] L. Rabiner, M. Cheng, A. Rosenberg, and C. MGonegal, "A comparative performance study of several pitch detection algorithms," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 5, pp. 399–418, 1976.

[556] DVSI, *Inmarsat-M Voice Codec*, Issue 3.0 ed., August 1991.

[557] M. Sambur, A. Rosenberg, L. Rabiner, and C. McGonegal, "On reducing the buzz in LPC synthesis," *Journal of the Acoustical Society of America*, vol. 63, pp. 918–924, March 1978.

[558] A. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," *Journal of the Acoustical Society of America*, vol. 49, no. 2 pt.2, pp. 583–590, 1971.

[559] T. Koornwinder, *Wavelets: An Elementary Treatment of Theory and Applications*. World Scientific, 1993.

[560] C. Chui, *Wavelet Analysis and its Applications*, vol. I: An Introduction to Wavelets. New York, USA: Academic Press, 1992.

[561] C. Chui, *Wavelet Analysis and its Applications*, vol. II: Wavelets: A Tutorial in Thoery and Applications. New York, USA: Academic Press, 1992.

[562] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, pp. 14–38, October 1991.

[563] A. Graps, "An introduction to wavelets," *IEEE Computational Science & Engineering*, pp. 50–61, Summer 1995.

[564] A. Cohen and J. Kovačević, "Wavelets: The mathematical background," *Proceedings of the IEEE*, vol. 84, pp. 514–522, April 1996.

[565] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, pp. 961–1005, September 1990.

[566] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, July 1989.

[567] H. Baher, *Analog & Digital Signal Processing*. New York, USA: John Wiley and Sons, 1990.

[568] J. Stegmann, G. Schröder, and K. Fischer, "Robust classification of speech based on the dyadic wavelet transform with application to CELP coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [616], pp. 546–549.

[569] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 710–732, July 1992.

[570] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proceedings of the IEEE*, vol. 84, pp. 626–638, April 1996.

[571] C. Li, C. Zheng, and C. Tai, "Detection of ECG characteristic points using wavelet transforms," *IEEE Transactions in Biomedical Engineering*, vol. 42, pp. 21–28, January 1995.

[572] S. Mallat and W. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, pp. 617–643, March 1992.

[573] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.

[574] R. Sukkar, J. LoCicero, and J. Picone, "Design and implementation of a robust pitch detector based on a parallel processing technique," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 441–451, February 1988.

[575] R. Steele and L. Hanzo, eds., *Mobile Radio Communications*. Piscataway, NJ, USA: IEEE Press, 1999.

[576] F. Brooks, B. Yeap, J. Woodard, and L. Hanzo, "A sixth-rate, 3.8kbps GSM-like speech transceiver," in *Proceeding of ACTS Mobile Communication Summit '98* [612], pp. 647–652.

[577] F. Brooks, E. Kuan, and L. Hanzo, "A 2.35kbps joint-detection CDMA speech transceiver," in *Proceeding of VTC'99 (Spring)* [613], pp. 2403–2407.

[578] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," in *Proceedings of the International Conference on Communications*, pp. 1009–1013, June 1995.

[579] P. Robertson, "Illuminating the structure of code and decoder of parallel concatenated recursive systematic (turbo) codes," *IEEE Globecom*, pp. 1298–1303, 1994.

[580] W. Koch and A. Baier, "Optimum and sub-optimum detection of coded data disturbed by time-varying inter-symbol interference," *IEEE Globecom*, pp. 1679–1684, December 1990.

[581] J. Erfanian, S. Pasupathy, and G. Gulak, "Reduced complexity symbol dectectors with parallel structures for ISI channels," *IEEE Transactions on Communications*, vol. 42, pp. 1661–1671, 1994.

[582] J. Hagenauer and P. Hoeher, "A viterbi algorithm with soft-decision outputs and its applications," in *IEEE Globecom*, pp. 1680–1686, 1989.

[583] C. Berrou, P. Adde, E. Angui, and S. Faudeil, "A low complexity soft-output viterbi decoder architecture," in *Proceedings of the International Conference on Communications*, pp. 737–740, May 1993.

[584] L. Rabiner, C. McGonegal, and D. Paul, *FIR Windowed Filter Design Program - WINDOW*, ch. 5.2. IEEE Press, 1979.

[585] S. Yeldner, A. Kondoz, and B. Evans, "Multiband linear predictive speech coding at very low bit rates," *IEE Proceedings in Vision, Image and Signal Processing*, vol. 141, pp. 284–296, October 1994.

[586] A. Klein, R. Pirhonen, J. Skoeld, and R. Suoranta, "FRAMES multiple access mode 1 — wideband TDMA with and without spreading," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'97* [624], pp. 37–41.

[587] J. Flanagan and R. Golden, "Phase vocoder," *The Bell System Technical Journal*, pp. 1493–1509, November 1966.

[588] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on sinusoidal representation," *IEEE Transactions on Acustics, Speech and Signal Processing*, vol. 34, pp. 744–754, August 1986.

[589] L. Almeida and J. Tribolet, "Nonstationary spectral modelling of voiced speech," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 31, pp. 664–677, June 1983.

[590] E. George and M. Smith, "Analysis-by-synthesis/overlap-add sinusoidal modelling applied to the analysis and synthesis of musical tones," *Journal of the Audio Engineering Society*, vol. 40, pp. 497–515, June 1992.

[591] E. George and M. Smith, "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," *IEEE Transaction on Speech and Audio Processing*, vol. 5, pp. 389–406, September 1997.

[592] R. McAulay and T. Quatieri, "Pitch estimation and voicing detection based on a sinusoidal speech model," in *Proceedings of ICASSP' 90*, pp. 249–252, 1990.

[593] R. McAulay and T. Quatieri, "Sinusoidal coding," in *Speech Coding and Synthesis* (W.B.Keijn and K.K.Paliwal, eds.), ch. 4, Netherlands: Elsevier Science, 1995.

[594] R. McAulay, T. Parks, T. Quatieri, and M. Sabin, "Sine-wave amplitude coding at low data rates," in *Advances in Speech Coding* (V. B.S.Atal and A.Gersho, eds.), pp. 203–214, Dordrecht: Kluwer Academic Publishers, 1991.

[595] M. Nishiguchi and J. Matsumoto, "Harmonic and noise coding of LPC residuals with classified vector quantization," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'95)* [614], pp. 484–487.

[596] V. Cuperman, P. Lupini, and B. Bhattacharya, "Spectral excitation coding of speech at 2.4kb/s," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'95)* [614], pp. 496–499.

[597] S. Yeldner, A. Kondoz, and B. Evans, "High quality multiband LPC coding of speech at 2.4kbit/s," *Electronics Letters*, vol. 27, no. 14, pp. 1287–1289, 1991.

[598] H. Yang, S.-N. Koh, and P. Sivaprakasapillai, "Pitch synchronous multi-band (PSMB) speech coding," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'95)* [614], pp. 516–518.

[599] E. Erzin, A. Kumar, and A. Gersho, "Natural quality variable-rate spectral speech coding below 3.0kbps," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)* [617], pp. 1579–1582.

[600] C. Papanastasiou and C. Xydeas, "Efficient mixed excitation models in LPC based prototype interpolation speech coders," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)* [617], pp. 1555–1558.

[601] O. Ghitza, "Auditory models and human performance in tasks related to speech coding and speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 2, pp. 115–132, January 1994.

[602] K. Kryter, "Methods for the calculation of the articulation index," tech. rep., American National Standards Institute, 1965.

[603] U. Halka and U. Heute, "A new approach to objective quality-measures based on attribute matching," *Speech Communications*, vol. 11, pp. 15–30, 1992.

[604] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *Journal on Selected Areas in Communications*, vol. 10, pp. 819–829, June 1992.

[605] T. Barnwell III and A. Bush, "Statistical correlation between objective and subjective measures for speech quality," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'78*, (Tulsa, Okla, USA), pp. 595–598, IEEE, 10–12 April 1978.

[606] T. Barnwell III, "Correlation analysis of subjective and objective measures for speech quality," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'80* [625], pp. 706–709.

[607] P. Breitkopf and T. Barnwell III, "Segmental preclassification for improved objective speech quality measures," in *IEEE Proceedings of International Conference on Acoustic Speech Signal Processing*, pp. 1101–1104, 1981.

[608] L. Hanzo and L. Hinsenkamp, "On the subjective and objective evaluation of speech codecs," *Budavox Telecommunications Review*, no. 2, pp. 6–9, 1987.

[609] K. Kryter, "Masking and speech communications in noise," in *The effects of Noise on Man*, ch. 2, New York, USA: Academic Press, 1970. ISBN: 9994669966.

[610] A. House, C. Williams, M. Hecker, and K. Kryter, "Articulation testing methods: Consonated differentiation with a closed-response set," *Journal Acoustic Soc. Am.*, pp. 158–166, January 1965.

[611] IEEE, *Proceeding of VTC'99 (Fall)*, (Amsterdam, Netherlands), 19–22 September 1999.

[612] ACTS, *Proceeding of ACTS Mobile Communication Summit '98*, (Rhodes, Greece), 8–11 June 1998.

[613] IEEE, *Proceeding of VTC'99 (Spring)*, (Houston, TX, USA), 16–20 May 1999.

[614] IEEE, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'95)*, (Detroit, MI, USA), 9–12 May 1995.

[615] IEEE, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'94)*, (Adelaide, Australia), 19–22 April 1994.

[616] IEEE, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)*, (Atlanta, GA, USA), 7–10 May 1996.

[617] IEEE, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)*, (Munich, Germany), 21–24 April 1997.

[618] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'77*, (Hartford, CT, USA), 9–11 May 1977.

[619] IEEE, *Proceedings of IEEE VTC '94*, (Stockholm, Sweden), 8–10 June 1994.

[620] IEEE, *Proceedings of IEEE Vehicular Technology Conference (VTC'98)*, (Ottawa, Canada), 18–21 May 1998.

[621] ACTS, *Proceeding of ACTS Mobile Communication Summit '97*, (Aalborg, Denmark), 7–10 October 1997.

[622] J. Gibson, ed., *The Mobile Communications Handbook*. Boca Raton, FL, USA: CRC Press and IEEE Press, 1996.

[623] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'90*, (Albuquerque, New Mexico, USA), 3–6 April 1990.

[624] IEEE, *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'97*, (Marina Congress Centre, Helsinki, Finland), 1–4 September 1997.

[625] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'80*, (Denver, CO, USA), 9–11 April 1980.

[626] W. Tuttlebee, ed., *Cordless Telecommunications in Europe: The Evolution of Personal Communications*. London: Springer-Verlag, 1990. ISBN 3540196331.

[627] IEE, *Proceedings of IEE Conference on Radio Receivers and Associated Systems (RRAS'95)*, (Bath, UK), 26–28 September 1995.

[628] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'84*, (San Diego, CA, USA), 19–21 March 1984.

[629] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'89*, (Glasgow, Scotland, UK), 23–26 May 1989.

[630] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'91*, (Toronto, Ontario, Canada), 14–17 May 1991.

[631] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'87*, (Dallas, TX, USA), 6–9 April 1987.

[632] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'88*, (New York, NY, USA), 11–14 April 1988.

[633] IEEE, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'93)*, (Minneapolis, MN, USA), 27–30 April 1993.

[634] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'83*, (Boston, MA, USA), 14–16 April 1983.

[635] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'92*, March 1992.

[636] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'82*, May 1982.

Author Index

Symbols

, Jr [509] 547
3GPP TS 26.173, [521] 552

A

Adde, P. [583] 683, 714
Adoul, J-P. [170] 189
Adoul, J-P. [191] 221
Adoul, J-P. [162] 176, 282, 398, 440–444, 446
Adoul, J-P. [225] 298, 302, 303, 307
Adoul, J-P. [163] 176, 442, 444–446, 487
Adoul, J-P. [213] 274, 386
Adoul, J-P. [160] 175, 189, 274, 307, 323, 350, 386, 394
Adoul, J-P. [224] 295, 297
Adoul, J-P. [223] 295, 297
Adoul, J-P. [294] 443, 446
Adoul, J.P. [168] 181, 188, 298, 303, 304, 307, 441
Adoul, J.P. [479] 524
Adoul, J.P. [166] 181
Adoul, J.P. [139] 138
Adoul, J.P. [228] 304, 307
Agrawal, D. [487] 528
Akagiri, K. [40] 5, 6, 490, 491, 498, 499, 502, 507–509, 557
Akagiri, K. [413] 493
Akansu, A. [436] 495
Alamouti, S. [50] . . 7, 491, 528, 529, 533, 534
Alcaim, A. [282] 404
Aldroubi, A. [570] 622
Almeida, L.B. [589] 719
Almeida, L.B. [198] 244
Alouini, M-S. [301] 445
Ananthapadmanabhan, A. [244] 320
Andrews, W. [528] 566, 569, 570
Andrews, W. [535] 569, 570
Angui, E. [583] 683, 714
Anisur Rahham, M.D. [179] 203
Antti Toskala, [491] 528, 529, 557
Appleby, D.G. [132] 129, 131
Appleby, D.G. [135] 134, 181
Arimochi, K. [298] 445
Asghar, S. [82] 60
Atal, B. [234] 320
Atal, B.S. [526] 562, 563, 687
Atal, B.S. [551] 594, 621
Atal, B.S. [92] 94
Atal, B.S. [9] 100, 155, 563
Atal, B.S. [131] 129, 131
Atal, B.S. [136] 134
Atal, B.S. [52] 561
Atal, B.S. [116] . 117, 129, 134, 135, 139, 209, 325, 589
Atal, B.S. [128] 129, 134–136
Atal, B.S. [543] 578
Atal, B.S. [16] . . 101, 175–177, 323, 370, 564
Atal, B.S. [365] 517
Atal, B.S. [175] 202
Atal, B.S. [96] 96
Atal, B.S. [99] 100
Atungsiri, S.A. [187] 210, 211

B

Baghbadrani, D.K. [167] 181, 441
Baher, H. [567] 619
Bahl, L. [503] 529
Bahl, L.R. [221] 289, 449, 683
Baier, A. [580] 683
Baier, P.W. [264] 335, 336
Barbulescu, A.S [220] 289
Barnwell, T.P. [537] 571

Barnwell, T.P. III [538] 571
Barnwell, T.P. III [531] 567, 571, 572, 610–614, 663, 695, 696, 759
Barton, S.K. [318] 447
Barton, S.K. [317] 447
Bauch, G. [486] 528
Baum, K.L. [326] 447
Bennett, W.R. [63] 29
Beritelli, F. [246] 321
Berrou, C. [583] 683, 714
Berrou, C. [216] . 289, 447, 449, 680, 681, 713
Berrou, C. [502] 528, 529, 534
Berrou, C. [217] . 289, 447, 449, 680, 681, 713
Besette, B. [225] 298, 302, 303, 307
Bessette, B. [520] 548, 549
Bessette, B. [358] 480
Bessette, B. [224] 295, 297
Bessette, B. [223] 295, 297
Bhattacharya, B. [596] 723, 724
Bielefeld, M.R. [547] 581
Bingham, J.A.C. [320] 447
Black, A.W. [161] . . . 176, 437, 438, 440, 445, 446, 487
Blocher, P. [257] 323, 328
Blumstein, S.E. [20] 562
Bogenfeld, E. [497] 528
Bogenfeld, E. [323] 447
Borowski, J. [497] 528
Borowski, J. [323] 447
Bosi, M. [40] . . . 5, 6, 490, 491, 498, 499, 502, 507–509, 557
Bosi, M. [38] 5, 6, 489, 491
Bosi, M. [412] 491, 501
Boudreaux-Bartels, G.F. [554] . 594, 621, 627, 631
Bradley, A.B. [442] 496, 501
Brandenburg, K. [40] 5, 6, 490, 491, 498, 499, 502, 507–509, 557
Brandenburg, K. [35] 5, 6, 489, 491, 496
Brandenburg, K. [419] 493
Brandenburg, K. [426] 494, 495
Brandenburg, K. [44] 5, 6, 490, 491, 517
Brandenburg, K. [445] 498
Brandenburg, K. [38] 5, 6, 489, 491
Brandenburg, K. [452] 507, 508
Brind'Amour, A. [529] 566–568, 723
Brock, D.P. [549] 581
Bruhn, S. [231] 319, 323
Bruhn, S. [29] 4
Bruhn, S. [257] 323, 328
Bryden, K. [529] 566–568, 723
Buzo, A. [280] 378, 440

C

Calderbank, A. [490] 528
Calderbank, A. [489] 528, 534
Calderbank, A. [485] 528
Calderbank, A. [483] 528, 533
Calderbank, A. [484] 528
Calderbank, A.R. [513] . . . 541, 542, 549, 552
Campbell, J.P. [186] 209, 210, 214, 242
Campbell, J.P. [197] 242, 245, 596
Campbell, W.M. [241] 320
Carciofy, C. [193] 230
Cattermole, K.W. [4] 32, 33, 561
Cellario, L. [237] 320
Champion, T. [534] . . 568, 569, 723, 724, 728
Champion, T.G. [546] 580
Chang, R.W. [310] 447
Chang, R.W. [492] 528
Charbonnier, A. [447] 502
Chaudhury, P. [481] 527
Cheetham, B.M.G. [121] 121
Cheetham, B.M.G. [125] 125
Chen, J-H. [275] 356
Chen, J-H. [272] 352
Chen, J-H. [274] 352, 354
Chen, J-H. [108] 102, 247, 352, 355, 366
Chen, J-H. [273] 352, 378
Chen, J-H. [276] 364, 378, 408, 440
Chen, J-H. [94] . . 96, 101, 102, 104, 349–352, 355, 357, 363, 401
Chen, J-H. [174] 201
Chen, J-H. [110] 103, 264, 268, 284, 327, 607, 609
Chen, J.H. [141] 138
Cheng, M.J. [555] 594
Cherriman, P. [500] 528
Cherriman, P.J. [514] 542, 548
Cheung, J.C.S [155] 171, 173, 222
Chiariglione, L. [30] 5, 489
Chiariglione, L. [31] 5, 489
Chiariglione, L. [32] 5, 489
Chow, P.S. [320] 447
Choy, E. [244] 320
Chu, C.C. [172] 197
Chua, S. [266] 336, 445
Chua, S. [307] 447
Chuang, J. [488] 528
Chuang, J.C.I. [498] 528

Chui, C.K. [560] 622, 623
Chui, C.K. [561] 619, 622
Chung, J.Y. [518] 547–549, 557, 558
Chung, J.Y. [516]543, 549, 552, 553, 557, 558
Cimini, J. [493] 528
Cimini, L.J. [311] 447
Cioffi, J.M. [320] 447
Classen, F. [315] 447
Classen, F. [316] 447
Cocke, J. [221] 289, 449, 683
Cocke, J. [503] 529
Cohen, A. [564] 619
Combescure, P. [143] 138, 456, 457
Combescure, P. [336] 463
Combescure, P. [170] 189
Contin, L. [45] 5, 490
Costello, D.J. [509] 547
Cox, R.V. [275] 356
Cox, R.V. [272] 352
Cox, R.V. [136] 134
Cox, R.V. [94] . . . 96, 101, 102, 104, 349–352, 355, 357, 363, 401
Cox, R.V. [174] 201
Cox, R.V. [2] 345
Cox, R.V. [185] 209, 214–216
Cox, R.V. [1] 345
Cox, R.V. [519] 548
Crochiere, R.E. [284] 419, 561, 687
Crochiere, R.E. [285] 419
Crochiere, R.E. [429] 494
Crochiere, R.E. [430] 494
Cuperman, V. [52] 561
Cuperman, V. [596] 723, 724
Cuperman, V. [102] 101
Cuperman, V. [238] 320

D

D'Agnoli, S.L.Q. [282] 404
Das, A. [232] 320
Das, A. [244] 320
Daubechies, I. [565] 619, 620
Davidson, G. [40] . . . 5, 6, 490, 491, 498, 499, 502, 507–509, 557
Davidson, G. [412] 491, 501
Davidson, G. [133] 131, 134, 320
Davis, M. [412] 491, 501
De Jaco, A. [244] 320
De Jaco, A. [235] 320, 321
de La Noue, P. [532] 567, 570, 571
De Marca, J.R.B. [277] 364, 409
De Marca, J.R.B. [282] 404
Deep Sen, [540] 573
Degroat, R.D. [180] 204
Deller, J.R. [19] 562
Delprat, M. [168] 181, 188, 298, 303, 304, 307, 441
Deprettere, E.F. [478] 521, 523, 524
Deprettere, E.F. [11] . . 157, 158, 160, 162, 564
Deprettere, E.F. [480] 524
Déry, S. [529] 566–568, 723
Di Benedetto, M.G. [319] 447
Dietz, M. [40] . . 5, 6, 490, 491, 498, 499, 502, 507–509, 557
Dietz, M. [455] 508
Dite, W. [65] 32
Doward, S. [414] 493
Dowling, E.M. [180] 204

E

Ebrahimi, T. [508] 541, 542
Edler, B. [45] 5, 490
Edler, B. [468] 516
Edler, B. [444] 498, 501
Edler, B. [49] 6, 491, 517
Edler, B. [440] 495
Ekudden, E. [231] 319, 323
Ekudden, E. [29] 4
El-Jaroudi, A. [552] 594, 621
Erdmann, C. [337] 464, 528
Erfanian, J.A. [581] 683
Eriksson, T. [148] 149
Eriksson, T. [149] 149, 150
Erzin, E. [599] 723
Esteban, D. [286] 421, 422, 424, 425, 692
Esteban, D. [428] 494, 495
Evans, B.G. [187] 210, 211
Evans, B.G. [161] . . . 176, 437, 438, 440, 445, 446, 487
Evans, B.G. [130] 129
Evans, B.G. [188] 212
Evans, B.G. [293] 439
Evans, B.G. [597] 723
Evans, B.G. [585] 693, 758

F

Faili, M. [409] 529, 531, 537
Failli, M. [331] . . 449, 681, 683, 684, 714, 715
Failli, M. [269] 336
Farvardin, N. [137] 135
Farvardin, N. [115] 117, 134

Fastl, H. [424] 494, 503, 516
Faudeil, S. [583] 683, 714
Fazel, K. [321] 447
Fazel, K. [312] 447
Fazel, K. [494] 528
Fazel, P.R.K. [495] 528
Ferreira, A. [448] 502
Fettweis, G. [312] 447
Fettweis, G. [494] 528
Fielder, L. [40] . 5, 6, 490, 491, 498, 499, 502, 507–509, 557
Fielder, L. [412] 491, 501
Fingscheidt, T. [523] 557
Fischer, K. [143] 138, 456, 457
Fischer, K. [336] 463
Fischer, K. [337] 464, 528
Fischer, K.A. [568] 621, 627, 631
Flanagan, J.L. [284] 419, 561, 687
Flanagan, J.L. [429] 494
Flanagan, J.L. [587] 719
Flannery, B.P. [177] 203–206, 404, 605
Fletcher, H. [421] 493
Fortune, P.M. [184] 209, 216, 217, 220
Foschini, G. Jr [482] 527
Fransen, L.J. [118] 118–121, 124, 577
Fratti, M. [176] 202, 207
Frullone, M. [193] 230
Fuchs, H. [40] . . 5, 6, 490, 491, 498, 499, 502, 507–509, 557
Fudseth, A. [138] 137, 138
Furui, S. [22] 16

G

Galand, C. [286] 421, 422, 424, 425, 692
Galand, C. [428] 494, 495
Galand, C.R. [289] 427
Gans, M. [482] 527
Gardner, W. [235] 320, 321
Gardner, W. [206] 253, 255, 261
Gardner, W. [48] 6, 491, 528
Geher, K. [122] 123
George, E.B. [590] . . 720, 721, 723, 732–734, 754, 758
George, E.B. [591] . . . 720, 721, 733, 754, 758
George, E.B. [537] 571
Gerrits, A.J. [476] 521
Gersho, A. [274] 352, 354
Gersho, A. [52] 561
Gersho, A. [108] 102, 247, 352, 355, 366
Gersho, A. [273] 352, 378
Gersho, A. [110] 103, 264, 268, 284, 327, 607, 609
Gersho, A. [232] 320
Gersho, A. [599] 723
Gersho, A. [236] 320
Gersho, A. [142] 138
Gersho, A. [126] 129, 142, 143, 364, 738–740
Gersho, A. [239] 320, 321
Gersho, A. [101] 101
Gersho, A. [133] 131, 134, 320
Gersho, A. [278] 364
Gerson, I.A. [204] 249, 251, 271, 272
Gerson, I.A. [202] 247, 249, 251, 269
Gerson, I.A. [164] 181, 187
Gerson, I.A. [203] 247, 249, 251, 269
Gerson, I.A. [211] 269, 272
Ghiselli-Crippa, T. [552] 594, 621
Ghitza, O. [601] 744, 745
Gish, H. [91] 78, 129, 132
Glavieux, A. [216] . . . 289, 447, 449, 680, 681, 713
Glavieux, A. [502] 528, 529, 534
Glavieux, A. [217] . . . 289, 447, 449, 680, 681, 713
Glisson, T.H. [68] 38
Golden, R.M. [587] 719
Goldsmith, A.J. [300] 445
Goldsmith, A.J. [266] 336, 445
Goldsmith, A.J. [307] 447
Goldsmith, A.J. [301] 445
Goldsmith, A.J. [302] 445
Golub, G.H. [178] 203, 204
Goodman, D.J. [192] 221, 226, 229, 235
Gordos, G. [15] 105, 107, 110, 111
Gray, A.H. [86] 67, 134
Gray, A.H. Jr [5] 16, 87
Gray, R. [233] 320
Gray, R.M. [280] 378, 440
Gray, R.M. [126] 129, 142, 143, 364, 738–740
Grazioso, P. [193] 230
Greenwood, D.D. [422] 493
Griffin, D.W. [103] . . 101, 566, 567, 569, 570, 687
Grill, B. [507] 541, 542
Gulak, G. [581] 683
Guyader, A.L. [336] 463

H

Haagen, J. [533] 567, 573, 653
Haavisto, P. [225] 298, 302, 303, 307

Haavisto, P. [228] 304, 307
Hagen, R. [540] 573
Hagenauer, J. [582] 683, 714
Hagenauer, J. [58] 15
Hagenauer, J. [27] 683
Hagenauer, J. [218] 289
Hagenauer, J. [519] 548
Hall, J.L. [543] 578
Hanauer, S.L. [526] 562, 563, 687
Hankanen, T. [225] 298, 302, 303, 307
Hansen, J.H.L. [19] 562
Harborg, H. [138] 137, 138
Harri Holma, [491] 528, 529, 557
Hashimoto, S. [173] 200, 201
Hassanein, H. [529] 566–568, 723
Hassanein, H. [102] 101
Hayashi, S. [212] 274
Haykin, S. [72] 45, 46
Hbner, J. [497] 528
Heddle, R. [413] 493
Hellman, R. [425] 494
Hellwig, K. [231] 319, 323
Hellwig, K. [29] 4
Hellwig, K. [257] 323, 328
Herre, J. [40] . . . 5, 6, 490, 491, 498, 499, 502, 507–509, 557
Herre, J. [451] 505
Herre, J. [452] 507, 508
Herre, J. [507] 541, 542
Herre, J. [433] 494, 506
Hess, W. [14] 593
Hikmet Sari, I.J. [496] 528
Hiotakakos, D.J. [541] 574, 575, 641, 646, 648, 652, 653, 659, 660, 684, 758, 807
Ho, P. [238] 320
Hoeher, P. [582] 683, 714
Hoeher, P. [578] 681, 683
Hoffmann, R. [13] 162
Holmes, J.N. [539] 572, 610, 612
Holmes, W.H. [95] 96
Holtzwarth, H. [64] 32
Honda, M. [85] 67, 210, 578
Honda, M. [464] 514, 515
Hong, C. [270] 351, 371, 382
Honkanen, T. [228] 304, 307
Huang, J.J.Y. [134] 131
Huber, J.B. [314] 447
Hübner, J. [323] 447
Huffman, D.A. [456] 510
Huges, P.M. [125] 125
Huggins, A.W.F. [536] 571, 743
Hwang, W.L. [572] 622

I

Iijima, K. [473] 517
Ikeda, K. [458] 513, 514, 516
Ikeda, K. [459] 513
Ikeda, K. [210] 261, 565
Ikeda, K. [460] 513
Ikeda, K. [463] 513, 515, 557
Ikedo, J. [210] 261, 565
Ireton, M.A. [165] 181
Ireton, M.A. [167] 181, 441
ISO/IEC JTC1/SC29/WG11 N2503, [506] 540, 549, 551
Itakura, F. [111] 105
Itakura, F. [112] 105
Itakura, F. [144] 139, 324, 577
Itakura, F. [113] 105
Itakura, F. [123] 125
Itakura, F. [173] 200, 201
Itoh, K. [113] 105
Itoh, K. [85] 67, 210, 578
Itoh, K. [87] 67
Iwakami, N. [458] 513, 514, 516
Iwakami, N. [459] 513
Iwakami, N. [450] 502, 513
Iwakami, N. [462] 513
Iwakami, N. [460] 513
Iwakami, N. [463] 513, 515, 557

J

Jacobs, P. [235] 320, 321
Jacobs, P. [206] 253, 255, 261
Jafarkhani, H. [485] 528
Jafarkhani, H. [484] 528
Jain, A.K. [69] 38, 53–55, 131, 132
Jarvinen, K. [520] 548, 549
Jarvinen, K. [225] 298, 302, 303, 307
Jarvinen, K. [228] 304, 307
Jasiuk, M.A. [202] 247, 249, 251, 269
Jasiuk, M.A. [164] 181, 187
Jasiuk, M.A. [203] 247, 249, 251, 269
Jasiuk, M.A. [211] 269, 272
Jayant, N. [94] . . . 96, 101, 102, 104, 349–352, 355, 357, 363, 401
Jayant, N.S. [277] 364, 409
Jayant, N.S. [106] 102, 352, 366
Jayant, N.S. [107] 102, 352, 366

Jayant, N.S. [272] 352
Jayant, N.S. [78] 50, 51, 53
Jayant, N.S. [417] 493
Jayant, N.S. [431] 494, 508, 510
Jayant, N.S. [10] 20, 23, 29, 30, 32, 36, 38, 39, 55, 56, 199, 435, 561, 605
Jeanclaude, I. [322] 447
Jelinek, F. [221] 289, 449, 683
Jelinek, F. [503] 529
Jelinek, M. [520] 548, 549
Jennings, A. [75] 45, 131, 132
Jin, A. [458] 513, 514, 516
Jin, A. [463] 513, 515, 557
Johansen, F. [138] 137, 138
Johnston, J. [451] 505
Johnston, J. [433] 494, 506
Johnston, J. [417] 493
Johnston, J. [414] 493
Johnston, J. [453] 507, 508
Johnston, J. [448] 502
Johnston, J. [427] 494, 495
Johnston, J.D. [287] 421, 427
Johnston, J.D. [291] 433
Jones, A.E. [318] 447
Juang, B-H. [117] . . . 117, 120–122, 124, 135, 139, 209
Juang, J. [244] 320
Jung, P. [219] 289, 290

K

Kabal, P. [119] 121, 123, 124
Kabal, P. [172] 197
Kabal, P. [295] 445
Kadambe, S. [554] 594, 621, 627, 631
Kai-Bor Yu, [179] 203
Kaiser, S. [321] 447
Kaiser, S. [495] 528
Kaleh, G.K. [264] 335, 336
Kalet, I. [324] 447
Kamio, Y. [297] 445
Kamio, Y. [305] 446
Kamio, Y. [299] 445
Kaneko, T. [462] 513
Kang, G.S. [118] 118–121, 124, 577
Kapanen, P. [225] 298, 302, 303, 307
Karam, G. [496] 528
Karam, G. [322] 447
Kataoka, A. [212] 274
Kataoka, A. [170] 189
Kawashima, T. [239] 320, 321
Keller, T. [214] 288, 292–294, 447
Keller, T. [308] 447, 450
Keller, T. [51] 7, 491, 528, 529, 534
Keller, T. [327] 447
Keller, T. [499] 528
Keller, T. [328] 447, 451
Keller, T. [248] 321, 322, 335, 336
Keller, T. [255] 322
Keller, T. [159] . 171, 221, 222, 235, 288, 410, 445
Keller, T. [512] 541, 542, 547, 549, 552
Keller, T. [254] 322
Ketchum, R.H. [199] 245
Ketchum, R.H. [281] 395
Kim, Y. [449] 502, 511
Kirchherr, R. [143] 138, 456, 457
Kirchherr, R. [336] 463
Kitawaki, N. [113] 105
Kitawaki, N. [85] 67, 210, 578
Kitawaki, N. [87] 67
Kitawaki, N. [462] 513
Kleider, J.E. [243] 320
Kleider, J.E. [241] 320
Kleijn, W.B. [185] 209, 214–216
Kleijn, W.B. [199] 245
Kleijn, W.B. [281] 395
Kleijn, W.B. [189] 215
Kleijn, W.B. [105] 102, 573, 646, 758
Kleijn, W.B. [533] 567, 573, 653
Kleijn, W.B. [540] 573
Kleijn, W.B. [335] 463, 507
Kleijn, W.B. [56] 323, 324, 561
Klein, A. [586] 713, 714
Klein, A. [264] 335, 336
Knudson, J. [138] 137, 138
Koch, W. [580] 683
Koenen, R. [504] 540
Koenen, R. [505] 540
Koenen, R. [42] 5, 490, 528
Koh, S-N [598] 723
Kohler, M.A. [527] 566
Kohler, M.A. [546] 580
Komaki, S. [249] 321, 336
Kondoz, A.M. [187] 210, 211
Kondoz, A.M. [161] . 176, 437, 438, 440, 445, 446, 487
Kondoz, A.M. [55] 143, 149, 176, 561
Kondoz, A.M. [130] 129
Kondoz, A.M. [188] 212
Kondoz, A.M. [293] 439

Kondoz, A.M. [597] 723
Kondoz, A.M. [585] 693, 758
Koora, K. [497] 528
Koora, K. [323] 447
Koornwinder, T.H. [559] 619, 622, 623
Kovać, J. [564] 619
Kovać, J. [573] 622
Kraisinsky, D.J. [199] 245
Krasinski, D.J. [281] 395
Kreamer, E.W. [550] 581
Kreamer, E.W. [548] 581
Krishna, H. [261] 332, 333
Krishna, H. [262] 332, 333
Kroon, P. [136] 134
Kroon, P. [185] 209, 214–216
Kroon, P. [1] 345
Kroon, P. [478] 521, 523, 524
Kroon, P. [170] 189
Kroon, P. [11] 157, 158, 160, 162, 564
Kroon, P. [234] 320
Kroon, P. [480] 524
Kuan, E.L. [577] 673
Kuan, E.L. [309] 447, 714
Kuan, E.L. [253] 322, 323, 328, 336
Kuan, E.L. [252] 322
Kull, B. [497] 528
Kull, B. [323] 447
Kumar, A. [599] 723
Kunz, O. [44] 5, 6, 490, 491, 517
Kwan Truong, [537] 571

L

Laflamme, C. [191] 221
Laflamme, C. [162] . . 176, 282, 398, 440–444, 446
Laflamme, C. [479] 524
Laflamme, C. [139] 138
Laflamme, C. [225] 298, 302, 303, 307
Laflamme, C. [228] 304, 307
Laflamme, C. [163] . . 176, 442, 444–446, 487
Laflamme, C. [213] 274, 386
Laflamme, C. [160] . . 175, 189, 274, 307, 323, 350, 386, 394
Laflamme, C. [224] 295, 297
Laflamme, C. [223] 295, 297
Laflamme, C. [294] 443, 446
Lamblin, C. [143] 138, 456, 457
Lamblin, C. [336] 463
Lamblin, C. [166] 181
Lamblin, C. [337] 464, 528
Laroia, R. [137] 135
Lau, V.K.N. [306] 446
Laurent, P.A. [532] 567, 570, 571
Law, H.B. [171] 191
Le Guyader, A. [143] 138, 456, 457
Leach, B. [528] 566, 569, 570
LeBlanc, W.P. [240] 320
Lederer, D. [452] 507, 508
Lee, C. [235] 320, 321
Lee, C. [206] 253, 255, 261
Lee, K.Y. [130] 129
Lee, W.C.Y. [195] 238
Lefebvre, R. [520] 548, 549
Lefebvre, R. [139] 138
Lefebvre, R. [358] 480
Lepschy, A. [124] 125
Levine, S. [469] 516, 517
Levine, S. [471] 517
Levinson, S. [145] 139
Li, Y. [488] 528
Li, Y. [325] 447
Li, Y.G. [498] 528
Lieberman, P. [20] 562
Liew, T.H. [303] 446
Liew, T.H. [511] . 541–543, 547, 549, 552, 553
Liew, T.H. [256] 322, 333
Liew, T.H. [304] 446
Liew, T.H. [251] 322, 446
Lim, J.S. [103] . . 101, 566, 567, 569, 570, 687
Lin, K-Y. [261] 332, 333
Lin, S. [509] 547
Lin, Y.-C. [275] 356
Lin, Y.C. [94] 96, 101, 102, 104, 349–352, 355, 357, 363, 401
Linde, Y. [280] 378, 440
Linden, J. [148] 149
Linden, J. [149] 149, 150
Lloyd, S.P. [60] 29, 36
Lloyd, S.P. [61] 29, 36
LoCicero, J.L. [574] 632
LoCicero, J.L. [542] . 574, 641, 647, 673, 758, 807
Lokhoff, G.C.P. [415] 493
Lombardo, A. [246] 321
Lupini, P. [596] 723, 724
Lupini, P. [102] 101

M

Mabilleau, P. [168] . . 181, 188, 298, 303, 304, 307, 441

Mabilleau, P. [162] . . 176, 282, 398, 440–444, 446
Mabilleau, P. [479] 524
Macleod, M.D. [306] 446
Mahieux, Y. [447] 502
Mahmoud, S.A. [240] 320
Maitre, X. [283] 415, 429
Makhoul, J. [77] 46, 47
Makhoul, J. [76] 46, 105
Makhoul, J. [536] 571, 743
Makhoul, J. [91] 78, 129, 132
Makhoul, J. [114] 113
Mallat, S. [566] 619, 622, 626
Mallat, S. [569] . 621, 622, 624, 626, 758, 804, 805
Mallat, S. [572] 622
Malvar, H.S. [333] 457, 459
Malvar, H.S. [437] 495, 496
Malvar, H.S. [438] 495
Malvar, H.S. [446] 501
Mandarini, P. [319] 447
Manjunath, S. [244] 320
Mano, K. [210] 261, 565
Markel, J.D. [86] 67, 134
Markel, J.D. [5] 16, 87
Marques, J.S. [198] 244
Massaloux, D. [143] 138, 456, 457
Massaloux, D. [336] 463
Massaloux, D. [166] 181
Massaloux, D. [160] . 175, 189, 274, 307, 323, 350, 386, 394
Massaloux, D. [337] 464, 528
Matsumoto, J. [104] 102
Matsumoto, J. [595] 723
Matsumoto, J. [472] 517
Matsumoto, J. [473] 517
Matsuoka, H. [305] 446
Max, J. [62] 29, 36, 38
May, T. [313] 447
McAulay, R.J. [588] 719–722, 754
McAulay, R.J. [592] 720
McAulay, R.J. [534] . . 568, 569, 723, 724, 728
McAulay, R.J. [594] 723, 725, 743, 744
McAulay, R.J. [530] . . 566, 568, 569, 723, 724
McAulay, R.J. [593] . . 720, 723, 743, 744, 754
McAulay, R.J. [242] 320
McCree, A. [537] 571
McCree, A.V. [538] 571
McCree, A.V. [531] . . 567, 571, 572, 610–614, 663, 695, 696, 759
McGonegal, C.A. [584] 693
McGonegal, C.A. [557] 610, 612
Meares, D. [45] 5, 490
Meine, N. [467] 516
Melchner, M.J. [94] 96, 101, 102, 104, 349–352, 355, 357, 363, 401
Meng, T.H.Y. [470] 516, 517
Mermelstein, P. [150] 149
Meyr, H. [315] 447
Meyr, H. [316] 447
MGonegal, C.A. [555] 594
Mian, G.A. [124] 125
Miani, G.A. [176] 202, 207
Miki, S. [458] 513, 514, 516
Miki, S. [450] 502, 513
Miki, S. [210] 261, 565
Miki, S. [460] 513
Miki, S. [463] 513, 515, 557
Miki, T. [209] 261
Miki, T. [208] 261
Mikkola, H. [520] 548, 549
Moncet, J.L. [172] 197
Morinaga, N. [249] 321, 336
Morinaga, N. [297] 445
Morinaga, N. [305] 446
Morinaga, N. [298] 445
Morinaga, N. [299] 445
Morissette, S. [168] . . 181, 188, 298, 303, 304, 307, 441
Morissette, S. [191] 221
Morissette, S. [162] . . 176, 282, 398, 440–444, 446
Morissette, S. [479] 524
Morissette, S. [166] 181
Moriya, T. [458] 513, 514, 516
Moriya, T. [459] 513
Moriya, T. [450] 502, 513
Moriya, T. [212] 274
Moriya, T. [462] 513
Moriya, T. [210] 261, 565
Moriya, T. [460] 513
Moriya, T. [463] 513, 515, 557
Moriya, T. [464] 514, 515
Moriya, T. [461] 513, 515
Muenster, M. [499] 528
Muller, J-M. [211] 269, 272
Müller, S.H. [314] 447
Münster, M. [255] 322
Murashima, A. [338] 464

N

Nagabucki, H. [87] 67
Naguib, A. [487] 528
Naguib, A. [490] 528
Naguib, A. [489] 528, 534
Naijoh, M. [299] 445
Nanda, S. [192] 221, 226, 229, 235
Nasshan, M. [219] 289, 290
Natvig, J.E. [151] 154, 162
Ng, S.X. [518] 547–549, 557, 558
Ng, S.X. [522] 556
Ng, S.X. [516] . . 543, 549, 552, 553, 557, 558
Ng, S.X. [512] 541, 542, 547, 549, 552
Niranjan, M. [182] 207
Nishiguchi, M. [104] 102
Nishiguchi, M. [595] 723
Nishiguchi, M. [472] 517
Nishiguchi, M. [473] 517
Noll, A.M. [553] 594
Noll, P. [431] 494, 508, 510
Noll, P. [10] . 20, 23, 29, 30, 32, 36, 38, 39, 55, 56, 199, 435, 561, 605
Noll, P. [67] 38
Noll, P. [88] 76
Noll, P. [416] 493
Noll, P. [37] 5, 6, 489, 491, 496
Noll, P. [432] 494
Nomura, T. [477] 521, 523, 524
Nowack, J.M. [211] 269, 272
Nussbaumer, H.J. [289] 427
Nussbaumer, H.J. [288] 427
Nussbaumer, H.J. [441] 495, 502

O

O'Neal, J. [90] 78
O'Shaughnessy, D. [17] 16, 562
Ochsner, H. [81] 60
Offer, E. [218] 289
Ohmuro, H. [210] 261, 565
Ohya, T. [209] 261
Ohya, T. [208] 261
Oikawa, Y. [40] . 5, 6, 490, 491, 498, 499, 502, 507–509, 557
Ojanpare., T. [215] 288, 290
Omologo, M. [120] 121
Ong, L.K. [188] 212
Ordentlich, E. [292] 437
Ozawa, K. [338] 464
Ozawa, K. [477] 521, 523, 524

P

Paez, M.D. [68] 38
Painter, T. [420] . 493, 495, 497, 499, 501, 503
Paksoy, E. [232] 320
Paksoy, E. [236] 320
Palazzo, S. [246] 321
Paliwal, K.K. [335] 463, 507
Paliwal, K.K. [56] 323, 324, 561
Paliwal, K.K. [116] . . 117, 129, 134, 135, 139, 209, 325, 589
Pan, D. [34] 5, 6, 489, 491, 502, 503, 505
Panter, P.F. [65] 32
Papanastasiou, C. [600] 724
Papke, L. [218] 289
Park, S. [449] 502, 511
Parks, T. [594] 723, 725, 743, 744
Pasupathy, S. [581] 683
Pattison, R.J. [243] 320
Paul, D. [584] 693
Paulus, J.W. [140] 138, 456
Pereira, F. [508] 541, 542
Petit, J. [447] 502
Phamdo, N. [137] 135
Picone, J.W. [574] 632
Picone, J.W. [542] . . . 574, 641, 647, 673, 758, 807
Pietrobon, S.S. [220] 289
Pirhonen, R. [586] 713, 714
Press, W.H. [177] 203–206, 404, 605
Princen, J.P. [442] 496, 501
Proakis, J.G. [19] 562
Proakis, J.G. [332] 451
Purnhagen, H. [468] 516
Purnhagen, H. [465] 516
Purnhagen, H. [467] 516
Purnhagen, H. [466] 516

Q

Quackenbush, S. [40] 5, 6, 490, 491, 498, 499, 502, 507–509, 557
quackenbush, S. [414] 493
Quackenbush, S.R. [290] 433, 434, 446
Quackenbush, S.R. [43] 5, 6, 490, 491
Quatieri, T.F. [588] 719–722, 754
Quatieri, T.F. [592] 720
Quatieri, T.F. [594] 723, 725, 743, 744
Quatieri, T.F. [530] . . . 566, 568, 569, 723, 724
Quatieri, T.F. [593] . . . 720, 723, 743, 744, 754
Quatieri, T.F. [242] 320
Quinquis, C. [143] 138, 456, 457

Quinquis, C. [337] 464, 528
Quinquisd, C. [336] 463

R

Rabiner, L. [145] 139
Rabiner, L.R. [551] 594, 621
Rabiner, L.R. [584] 693
Rabiner, L.R. [555] 594
Rabiner, L.R. [6]45–47, 50, 87, 105, 111, 113, 575–577
Rabiner, L.R. [557] 610, 612
Ragot, S. [358] 480
Ramachandran, R.P. [119] 121, 123, 124
Ramachandran, R.P. [128] 129, 134–136
Ramamoorthy, V. [106] 102, 352, 366
Ramamoorthy, V. [107] 102, 352, 366
Rao, K.R. [334] 459
Raviv, J. [221] 289, 449, 683
Raviv, J. [503] 529
Remde, J.R. [9] 100, 155, 563
Riccardi, G. [176] 202, 207
Rioul, O. [562] 619–621
Riva, G. [193] 230
Robertson, P. [321] 447
Robertson, P. [579] 682
Robertson, P. [578] 681, 683
Rohling, H. [313] 447
Rosenberg, A.E. [555] 594
Rosenberg, A.E. [558] 610–612
Rosenberg, A.E. [557] 610, 612
Rothweiler, J.H. [439] 495, 501
Rotola-Pukkila, J. [520] 548, 549
Roucos, S. [91] 78, 129, 132
Roy, G. [295] 445
Ruf, M. [495] 528
Ruf, M.J. [321] 447

S

Sabin, M. [594] 723, 725, 743, 744
Saito, S. [111] 105
Saito, S. [112] 105
Salami, R. [520] 548, 549
Salami, R.A. [154] 168
Salami, R.A. [184] 209, 216, 217, 220
Salami, R.A. [162] . . 176, 282, 398, 440–444, 446
Salami, R.A. [479] 524
Salami, R.A. [139] 138
Salami, R.A. [225] 298, 302, 303, 307
Salami, R.A. [228] 304, 307
Salami, R.A. [132] 129, 131
Salami, R.A. [135] 134, 181
Salami, R.A. [163] . . . 176, 442, 444–446, 487
Salami, R.A. [213] 274, 386
Salami, R.A. [160] . . 175, 189, 274, 307, 323, 350, 386, 394
Salami, R.A. [224] 295, 297
Salami, R.A. [223] 295, 297
Salami, R.A. [367] 507, 524
Salami, R.A. [70] 45, 94, 96–99, 156–158, 160, 162, 176, 179, 182, 210, 214
Salami, R.A. [294] 443, 446
Salami, R.A. [71] 45–47, 50, 96–99, 113, 155–160, 162, 176, 179, 181, 182, 191, 421, 442
Salami, R.A. [153] 168
Sambur, M.R. [557] 610, 612
Sampei, S. [249] 321, 336
Sampei, S. [297] 445
Sampei, S. [305] 446
Sampei, S. [298] 445
Sampei, S. [299] 445
Sanchez-Calle, V.E. [294] 443, 446
Sari, H. [322] 447
Sasaoka, H. [297] 445
Schafer, R.W. [6] 45–47, 50, 87, 105, 111, 113, 575–577
Scharf, B. [423] 493, 508
Scheirer, E. [48] 6, 491, 528
Scheirer, E.D. [47] 6, 491
Schembra, G. [246] 321
Schmidt-Nielsen, A. [549] 581
Schnitzler, J. [143] 138, 456, 457
Schnitzler, J. [336] 463
Schnitzler, J. [140] 138, 456
Schnitzler, J. [337] 464, 528
Schreiner, P. [45] 5, 490
Schröder, G. [568] 621, 627, 631
Schroeder, M.R. [92] 94
Schroeder, M.R. [543] 578
Schroeder, M.R. [16] 101, 175–177, 323, 370, 564
Schroeder, M.R. [365] 517
Schultheis, P.M. [134] 131
Schur, J. [152] 164
Schwartz, R. [536] 571, 743
Sen, D. [95] 96
Seo, Y. [449] 502, 511
Sereno, D. [237] 320
Serizawa, M. [338] 464

Serizawa, M. [477] 521, 523, 524
Seshadri, N. [487] 528
Seshadri, N. [519] 548
Seshadri, N. [490] 528
Seshadri, N. [489] 528, 534
Seshadri, N. [128] 129, 134–136
Seshadri, N. [513] 541, 542, 549, 552
Seshadri, N. [483] 528, 533
Seymour, R.A. [171] 191
Shannon, C.E. [57] 15
Sharma, V. [239] 320, 321
Shepherd, S.J. [317] 447
Shimoyoshi, O. [413] 493
Shinobu Ono, [104] 102
Shlien, S. [36] 5, 6, 489, 491, 504
Shlien, S. [443] 497, 498
Shoham, Y. [292] 437
Shoham, Y. [127] 129
Shoham, Y. [200] 245
Singhal, S. [175] 202
Singhal, S. [96] 96
Singhal, S. [99] 100
Sinha, D. [414] 493
Sivaprakasapillai, P [598] 723
Sjoberg, J. [257] 323, 328
Skoeld, J. [586] 713, 714
Skoglung, J. [148] 149
Skoglung, J. [149] 149, 150
Sluijter, R.J. [476] 521
Sluyter, R.J. [11] 157, 158, 160, 162, 564
Sluyter, R.J. [480] 524
Sluyter, R.J. [12] 162
Smith (Eds), M.T.J. [436] 495
Smith, B. [66] 32
Smith, J.O. [469] 516, 517
Smith, J.O. III [471] 517
Smith, M.J.T. [590] . . 720, 721, 723, 732–734, 754, 758
Smith, M.J.T. [591] . . 720, 721, 733, 754, 758
So, K.K.M. [129] 129
Sofranek, R. [417] 493
Soheili, R. [293] 439
Sollenberger, N. [488] 528
Sollenberger, N.R. [498] 528
Sollenberger, N.R. [325] 447
Somerville, F.C.A. [515] 542, 550
Sondhi, M. [145] 139
Sondhi, M.M. [128] 129, 134–136
Sonohara, M. [413] 493
Soong, F.K. [117] . . . 117, 120–122, 124, 135, 139, 209
Spanias, A. [420] 493, 495, 497, 499, 501, 503
Sreenivas, T.V. [455] 508
Srinivasan, K. [236] 320
Steedman, R.A.J. [79] 60
Steele, R. [265] 336, 445
Steele, R. [158] . 171, 221, 224, 324, 332, 410, 680–683
Steele, R. [575] 648
Steele, R. [154] 168
Steele, R. [184] 209, 216, 217, 220
Steele, R. [296] 445
Steele, R. [71] 45–47, 50, 96–99, 113, 155–160, 162, 176, 179, 181, 182, 191, 421, 442
Steele, R. [3] 561
Steele, R. [339] 464, 522, 528
Steele, R. [153] 168
Steele, R. [155] 171, 173, 222
Steele, R. [194] 235–240
Stefanov, J. [98] 100, 117, 154, 162, 187, 221, 223, 224, 228, 229, 231, 235
Stegmann, J. [143] 138, 456, 457
Stegmann, J. [336] 463
Stegmann, J. [337] 464, 528
Stegmann, J. [568] 621, 627, 631
Stoll, G. [35] 5, 6, 489, 491, 496
Stoll, G. [445] 498
Street, J. [514] 542, 548
Su, H.Y. [191] 221
Suda, H. [209] 261
Suda, H. [208] 261
Suen, A.N. [201] 247
Sugamura, N. [123] 125
Sugamura, N. [115] 117, 134
Sugiyama, A. [44] 5, 6, 490, 491, 517
Sukkar, R.A. [574] 632
Sukkar, R.A. [542] . . 574, 641, 647, 673, 758, 807
Sun, J-D. [261] 332, 333
Sun, J-D. [262] 332, 333
Sundberg, C-E.W. [519] 548
Suoranta, R. [586] 713, 714
Supplee, L.M. [547] 581
Supplee, L.M. [527] 566
Suzuki, H. [413] 493
Szabo, N.S. [259] 332, 333

T

Takacs, GY. [15] 105, 107, 110, 111
Tanaka, R.I. [259] 332, 333
Taniguchi, T. [233] 320
Taori, R. [476] 521
Tardelli, J.D. [550] 581
Tardelli, J.D. [548] 581
Tarokh, V. [487] 528
Tarokh, V. [490] 528
Tarokh, V. [485] 528
Tarokh, V. [513] 541, 542, 549, 552
Tarokh, V. [483] 528, 533
Tarokh, V. [484] 528
Taylor, F.J. [260] 332
Teague, K.A. [528] 566, 569, 570
Teague, K.A. [535] 569, 570
Temerinac, M. [440] 495
Teukolsky, S.A. [177] 203–206, 404, 605
Thitimajshima, P. [216] ... 289, 447, 449, 680, 681, 713
Thitimajshima, P. [502] 528, 529, 534
Thorpe, T. [89] 77
Timor, U. [192] 221, 226, 229, 235
Tobias, J.V. [8] 434
Todd, C. [412] 491, 501
Tohkura, Y. [173] 200, 201
Torrance, J.M. [267] 336
Trancoso, I.M. [198] 244
Tremain, T. [186] 209, 210, 214, 242
Tremain, T.E. [197] 242, 245, 596
Tremain, T.E. [527] 566
Tremain, T.E. [196] 242, 562, 564, 566
Tremain, T.E. [546] 580
Tribolet, J.M. [589] 719
Tribolet, J.M. [198] 244
Tribolet, J.M. [430] 494
Tsutsui, K. [413] 493
Tzeng, F.F. [181] 207

U

Ubale, A. [142] 138
Unagami, S. [233] 320
Ungerb, $G.1001[510]$ $541, 542, 547, 549, 552$
Unser, M. [570] 622

V

Vaidyanathan, P.P. [435] 495, 496
Vaidyanathan, P.P. [434] 495
Vainio, J. [520] 548, 549
Vainio, J. [225] 298, 302, 303, 307
Vainio, J. [228] 304, 307
van der Waal, R.G. [454] 508
van Eetvelt, P.W.J. [317] 447
Van Loan, C.F. [178] 203, 204
Varaiya, P.P. [302] 445
Vary, P. [143] 138, 456, 457
Vary, P. [336] 463
Vary, P. [337] 464, 528
Vary, P. [523] 557
Vary, P. [12] 162
Vary, P. [13] 162
Veldhuis, R.N.J. [454] 508
Vercoe, B. [48] 6, 491, 528
Verma, T. [469] 516, 517
Verma, T.S. [470] 516, 517
Vernon, S. [412] 491, 501
Vetterli, M. [562] 619–621
Vetterli, M. [573] 622
Vetterling, W.T. [177] 203–206, 404, 605
Viaro, U. [124] 125
Villebrun, E. [578] 681, 683
Viswanathan, R. [536] 571, 743
Viswanathan, R. [114] 113
Viswanathan, V. [537] 571
Viterbi, A.J. [59] 15
Voiers, W.D. [544] 580
Voiers, W.D. [545] 580
Vook, F.W. [326] 447

W

Wakatsuki, R. [104] 102
Walls, B. [535] 569, 570
Wand, J.F. [201] 247
Wang, D. [141] 138
Wang, H-S. [271] 351
Wang, S. [101] 101
Wassell, I. [71] 45–47, 50, 96–99, 113, 155–160, 162, 176, 179, 181, 182, 191, 421, 442
Webb, W. [154] 168
Webb, W. [153] 168
Webb, W.T. [265] 336, 445
Webb, W.T. [51] 7, 491, 528, 529, 534
Webb, W.T. [248] 321, 322, 335, 336
Webb, W.T. [296] 445
Webb, W.T. [159] ... 171, 221, 222, 235, 288, 410, 445
Webb, W.T. [512] 541, 542, 547, 549, 552
Webb, W.T. [73] .. 45, 172, 228, 235, 288, 410
Webber, S.A. [284] 419, 561, 687

Webber, S.A. [429] 494
Welch, V. [186] 209, 210, 214, 242
Welch, V.C. [197] 242, 245, 596
Wilkinson, T.A. [318] 447
Williams, J.E.B. [155] 171, 173, 222
Williams, J.E.B. [194] 235–240
Winter, E.H. [211] 269, 272
Wong, C.H. [250] 322
Wong, C.H. [268] 336, 446
Wong, C.H. [309] 447, 714
Wong, C.H. [253] 322, 323, 328, 336
Wong, C.H. [500] 528
Wong, C.H. [303] 446
Wong, C.H. [304] 446
Wong, K.H.H. [74] 45, 47
Wong, K.H.J. [71] 45–47, 50, 96–99, 113, 155–160, 162, 176, 179, 181, 182, 191, 421, 442
Wyatt–Millington, C.W. [317] 447

X

Xydeas, C.S. [541] . . 574, 575, 641, 646, 648, 652, 653, 659, 660, 684, 758, 807
Xydeas, C.S. [165] 181
Xydeas, C.S. [600] 724
Xydeas, C.S. [167] 181, 441
Xydeas, C.S. [129] 129

Y

Yair Shoham, [540] 573
Yang, H. [598] 723
Yang, L-L. [263] 333
Yang, L.L. [256] 322, 333
Yao, T.C. [201] 247
Yeap, B.L. [576] 673
Yeap, B.L. [511] 541–543, 547, 549, 552, 553
Yee, M.S. [251] 322, 446
Yeldner, S. [597] 723
Yeldner, S. [585] 693, 758
Yip, P. [334] 459
Yong, M. [133] 131, 134, 320
Yuen, E. [238] 320

Z

Zarrinkoub, H. [150] 149
Zeger, K.A. [278] 364
Zehavi, E. [517] 547
Zeiberg, S. [497] 528
Zeisberg, S. [323] 447
Zelinski, R. [67] 38
Zelinski, R. [88] 76
Zelinski, R. [432] 494
Zhang, J. [271] 351
Zhong, S. [569] . 621, 622, 624, 626, 758, 804, 805
Zwicker, E. [418] 493
Zwicker, E. [424] 494, 503, 516