

Voice Compression and Communications:
Principles and Applications for Fixed and Wireless
Channels

by

©L. Hanzo, F.C.A. Somerville, J.P. Woodard
Department of Electronics and Computer Science,
University of Southampton, UK

Contents

Preface and Motivation	3
Acknowledgements	9
I Transmission Issues	11
1 The Propagation Environment	13
1.1 Introduction to Communications Issues	13
1.2 AWGN Channel	14
1.2.1 Background	14
1.2.2 Practical Gaussian Channels	15
1.2.3 Gaussian Noise	16
1.2.4 Shannon-Hartley Law	18
1.3 The Cellular Concept	19
1.4 Radio Wave Propagation	22
1.4.1 Background	22
1.4.2 Narrow-band fading Channels	24
1.4.3 Propagation Pathloss Law	25
1.4.4 Slow Fading Statistics	27
1.4.5 Fast Fading Statistics	28
1.4.6 Doppler Spectrum	33
1.4.7 Simulation of Narrowband Channels	35
1.4.7.1 Frequency-domain fading simulation	36
1.4.7.2 Time-domain fading simulation	37
1.4.7.3 Box-Müller Algorithm of AWGN generation	37
1.4.8 Wideband Channels	38
1.4.8.1 Modelling of Wideband Channels	38
1.5 Shannon's Message for Wireless Channels	43

2	Modulation and Transmission	47
2.1	The Wireless Communications Scene	47
2.2	Modulation Issues	49
2.2.1	Choice of Modulation	49
2.2.2	Quadrature Amplitude Modulation [47]	51
2.2.2.1	QAM overview	51
2.2.2.2	Modem Schematic	51
2.2.2.2.1	Gray Mapping and Phasor Constellation	51
2.2.2.2.2	Nyquist Filtering	54
2.2.2.2.3	Modulation and Demodulation	56
2.2.2.2.4	Data recovery	57
2.2.2.3	QAM Constellations	58
2.2.2.4	16QAM BER versus SNR Performance over AWGN Channels	61
2.2.2.4.1	Decision Theory	61
2.2.2.4.2	QAM Modulation and Transmission	63
2.2.2.4.3	16-QAM Demodulation	64
2.2.2.5	Reference Assisted Coherent QAM for Fading Channels	67
2.2.2.5.1	PSAM System Description	67
2.2.2.5.2	Channel Gain Estimation in PSAM	69
2.2.2.5.3	PSAM performance	71
2.2.2.6	Differentially detected QAM	72
2.2.3	Adaptive Modulation	76
2.2.3.1	Background to Adaptive Modulation	76
2.2.3.2	Optimisation of Adaptive Modems	79
2.2.3.3	Adaptive Modulation Performance	81
2.2.3.4	Equalisation Techniques	83
2.2.4	Orthogonal Frequency Division Multiplexing	84
2.3	Packet Reservation Multiple Access	86
2.4	Flexible Transceiver Architecture	89
3	Convolutional Channel Coding	93
3.1	Brief Channel Coding History	93
3.2	Convolutional Encoding	94
3.3	State and Trellis Transitions	96
3.4	The Viterbi Algorithm	98
3.4.1	Error-free hard-decision Viterbi decoding	98
3.4.2	Erroneous hard-decision Viterbi decoding	101
3.4.3	Error-free soft-decision Viterbi decoding	104
4	Block-based Channel Coding	107
4.1	Introduction	107
4.2	Finite Fields	108
4.2.1	Definitions	108
4.2.2	Galois Field Construction	111
4.2.3	Galois Field Arithmetic	113

CONTENTS

iii

4.3	RS and BCH Codes	114
4.3.1	Definitions	114
4.3.2	RS Encoding	116
4.3.3	RS Encoding Example	118
4.3.4	Circuits for Cyclic Encoders	122
4.3.4.1	Polynomial Multiplication	122
4.3.4.2	Shift Register Encoding Example	123
4.3.5	RS Decoding	126
4.3.5.1	Formulation of the RS Key-Equations	126
4.3.5.2	Peterson-Gorenstein-Zierler Decoder	131
4.3.5.3	PGZ Decoding Example	133
4.3.5.4	Berlekamp-Massey algorithm	138
4.3.5.5	Berlekamp-Massey Decoding Example	145
4.3.5.6	Forney Algorithm	149
4.3.5.7	Forney Algorithm Example	153
4.3.5.8	Error Evaluator Polynomial Computation	154
4.4	RS and BCH Codec Performance	157
4.5	Summary and Conclusions	160

II Speech Signals and Waveform Coding 163

5	Speech Signals and Coding	165
5.1	Motivation of Speech Compression	165
5.2	Basic Characterisation of Speech Signals	166
5.3	Classification of Speech Codecs	170
5.3.1	Waveform Coding	171
5.3.1.1	Time-domain Waveform Coding	171
5.3.1.2	Frequency-domain Waveform Coding	172
5.3.2	Vocoders	172
5.3.3	Hybrid Coding	173
5.4	Waveform Coding	174
5.4.1	Digitisation of Speech	174
5.4.2	Quantisation Characteristics	175
5.4.3	Quantisation Noise and Rate-Distortion Theory	176
5.4.4	Non-uniform Quantisation for a Known PDF: Companding	179
5.4.5	PDF-independent Quantisation by Logarithmic Compression	181
5.4.5.1	The μ -Law Compander	183
5.4.5.2	The A-law Compander	184
5.4.6	Optimum Non-uniform Quantisation	186
6	Predictive Coding	193
6.1	Forward Predictive Coding	193
6.2	DPCM Codec Schematic	194
6.3	Predictor Design	195
6.3.1	Problem Formulation	195

6.3.2	Covariance Coefficient Computation	197
6.3.3	Predictor Coefficient Computation	198
6.4	Adaptive One-word-memory Quantization	203
6.5	DPCM Performance	205
6.6	Backward-Adaptive Prediction	207
6.6.1	Background	207
6.6.2	Stochastic Model Processes	209
6.7	The 32 kbps G.721 ADPCM Codec	212
6.7.1	Functional G.721 Description	212
6.7.2	Adaptive Quantiser	213
6.7.3	G.721 Quantiser Scale Factor Adaptation	215
6.7.4	G.721 Adaptation Speed Control	215
6.7.5	G.721 Adaptive Prediction and Signal Reconstruction	217
6.8	Speech Quality Evaluation	219
6.9	G.726 and G.727 ADPCM Coding	220
6.9.1	Motivation	220
6.9.2	Embedded ADPCM coding	220
6.9.3	Performance of the Embedded G.727 ADPCM Codec	222
6.10	Rate-Distortion in Predictive Coding	225

III Analysis by Synthesis Coding 235

7 Analysis-by-synthesis Principles 237

7.1	Motivation	237
7.2	Analysis-by-synthesis Codec Structure	238
7.3	The Short-term Synthesis Filter	240
7.4	Long-Term Prediction	242
7.4.1	Open-loop Optimisation of LTP parameters	242
7.4.2	Closed-loop Optimisation of LTP parameters	248
7.5	Excitation Models	252
7.6	Adaptive Postfiltering	254
7.7	Lattice-based Linear Prediction	257

8 Speech Spectral Quantization 265

8.1	Log-area Ratios	265
8.2	Line Spectral Frequencies	269
8.2.1	Derivation of Line Spectral Frequencies	269
8.2.2	Determination of Line Spectral Frequencies	273
8.2.3	Chebyshev-description of Line Spectral Frequencies	275
8.3	Spectral Vector Quantization	281
8.3.1	Background	281
8.3.2	Speaker-adaptive Vector Quantisation of LSFs	281
8.3.3	Stochastic VQ of LPC Parameters	283
8.3.3.1	Background	283
8.3.3.2	The Stochastic VQ Algorithm	284

CONTENTS

v

8.3.4	Robust Vector Quantisation Schemes for LSFs	287
8.3.5	LSF Vector-quantisers in Standard Codecs	288
8.4	Spectral Quantizers for Wideband Speech Coding	290
8.4.1	Introduction to Wideband Spectral Quantisation	290
8.4.1.1	Statistical Properties of Wideband LSFs	291
8.4.1.2	Speech Codec Specifications	294
8.4.2	Wideband LSF Vector Quantizers	295
8.4.2.1	Memoryless Vector Quantization	295
8.4.2.2	Predictive Vector Quantization	300
8.4.2.3	Multimode Vector Quantization	301
8.4.3	Simulation Results and Subjective Evaluations	304
8.4.4	Conclusions on Wideband Spectral Quantisation	306
9	RPE Coding	309
9.1	Theoretical Background	309
9.2	The RPE-LTP GSM Speech encoder	316
9.2.1	Pre-processing	317
9.2.2	STP analysis filtering	317
9.2.3	LTP analysis filtering	319
9.2.4	Regular Excitation Pulse Computation	320
9.3	The RPE-LTP Speech Decoder	320
9.4	Bit-sensitivity of the GSM Codec	323
9.5	A 'Tool-box' Based Speech Transceiver	326
10	Forward-Adaptive CELP Coding	329
10.1	Background	329
10.2	The Original CELP Approach	331
10.3	Fixed Codebook Search	333
10.4	CELP Excitation Models	336
10.4.1	Binary Pulse Excitation	336
10.4.2	Transformed Binary Pulse Excitation	337
10.4.2.1	Excitation Generation	337
10.4.2.2	TBPE Bit Sensitivity	339
10.4.3	Dual-rate Algebraic CELP Coding	342
10.4.3.1	ACELP Codebook Structure	342
10.4.3.2	Dual-rate ACELP Bitallocation	344
10.4.3.3	Dual-rate ACELP Codec Performance	345
10.5	CELP Optimization	346
10.5.1	Introduction	346
10.5.2	Calculation of the Excitation Parameters	347
10.5.2.1	Full Codebook Search Theory	347
10.5.2.2	Sequential Search Procedure	349
10.5.2.3	Full Search Procedure	350
10.5.2.4	Sub-Optimal Search Procedures	352
10.5.2.5	Quantization of the Codebook Gains	353
10.5.3	Calculation of the Synthesis Filter Parameters	356

10.5.3.1	Bandwidth Expansion	356
10.5.3.2	Least Squares Techniques	357
10.5.3.3	Optimization via Powell's Method	360
10.5.3.4	Simulated Annealing and the Effects of Quantization	361
10.6	CELP Error-sensitivity	364
10.6.1	Introduction	364
10.6.2	Improving the Spectral Information Error Sensitivity	365
10.6.2.1	LSF Ordering Policies	365
10.6.2.2	The Effect of FEC on the Spectral Parameters	367
10.6.2.3	The Effect of Interpolation	368
10.6.3	Improving the Error Sensitivity of the Excitation Parameters	369
10.6.3.1	The Fixed Codebook Index	370
10.6.3.2	The Fixed Codebook Gain	370
10.6.3.3	Adaptive Codebook Delay	371
10.6.3.4	Adaptive Codebook Gain	372
10.6.4	Matching Channel Codecs to the Speech Codec	372
10.6.5	Error Resilience Conclusions	377
10.7	Dual-mode Speech Transceiver	378
10.7.1	The Transceiver Scheme	378
10.7.2	Re-configurable Modulation	378
10.7.3	Source-matched Error Protection	381
10.7.3.1	Low-quality 3.1 kBd Mode	381
10.7.3.2	High-quality 3.1 kBd Mode	385
10.7.4	Packet Reservation Multiple Access	386
10.7.5	3.1 kBd System Performance	388
10.7.6	3.1 kBd System Summary	391
10.8	Multi-slot PRMA Transceiver	392
10.8.1	Background and Motivation	392
10.8.2	PRMA-assisted Multi-slot Adaptive Modulation	393
10.8.3	Adaptive GSM-like Schemes	394
10.8.4	Adaptive DECT-like Schemes	396
10.8.5	Summary of Adaptive Multi-slot PRMA	397
11	Standard CELP Codecs	399
11.1	Background	399
11.2	The US DoD FS-1016 4.8 kbits/s CELP Codec	400
11.2.1	Introduction	400
11.2.2	LPC Analysis and Quantization	402
11.2.3	The Adaptive Codebook	402
11.2.4	The Fixed Codebook	403
11.2.5	Error Concealment Techniques	404
11.2.6	Decoder Post-Filtering	405
11.2.7	Conclusion	405
11.3	The IS-54 DAMPS speech codec	406
11.4	The JDC speech codec	409
11.5	The Qualcomm Variable Rate CELP Codec	412

CONTENTS

vii

11.5.1	Introduction	412
11.5.2	Codec Schematic and Bit Allocation	413
11.5.3	Codec Rate Selection	414
11.5.4	LPC Analysis and Quantization	414
11.5.5	The Pitch Filter	416
11.5.6	The Fixed Codebook	417
11.5.7	Rate 1/8 Filter Excitation	418
11.5.8	Decoder Post-Filtering	419
11.5.9	Error Protection and Concealment Techniques	419
11.5.10	Conclusion	420
11.6	Japanese Half-Rate Speech Codec	420
11.6.1	Introduction	420
11.6.2	Codec Schematic and Bit Allocation	421
11.6.3	Encoder Pre Processing	423
11.6.4	LPC Analysis and Quantization	423
11.6.5	The Weighting Filter	424
11.6.6	Excitation Vector 1	425
11.6.7	Excitation Vector 2	426
11.6.8	Quantization of the Gains	428
11.6.9	Channel Coding	429
11.6.10	Decoder Post Processing	431
11.7	The half-rate GSM codec	432
11.7.1	Half-rate GSM codec outline	432
11.7.2	Half-rate GSM Codec's Spectral Quantisation	434
11.7.3	Error protection	435
11.8	The 8 kbits/s G.729 Codec	436
11.8.1	Introduction	436
11.8.2	Codec Schematic and Bit Allocation	437
11.8.3	Encoder Pre-Processing	438
11.8.4	LPC Analysis and Quantization	439
11.8.5	The Weighting Filter	441
11.8.6	The Adaptive Codebook	442
11.8.7	The Fixed Algebraic Codebook	443
11.8.8	Quantization of the Gains	446
11.8.9	Decoder Post Processing	447
11.8.10	G.729 Error Concealment Techniques	449
11.8.11	G.729 Bit-sensitivity	450
11.8.12	Turbo-coded OFDM G.729 Speech Transceiver	451
11.8.12.1	Background	451
11.8.12.2	System Overview	452
11.8.12.3	Turbo Channel Encoding	452
11.8.12.4	OFDM in the FRAMES Speech/Data Sub-Burst	454
11.8.12.5	Channel model	455
11.8.12.6	Turbo-coded G.729 OFDM Parameters	456
11.8.12.7	Turbo-coded G.729 OFDM Performance	456
11.8.12.8	Turbo-coded G.729 OFDM Summary	457

11.8.13 G.729 Summary	457
11.9 The Reduced Complexity G.729 Annex A Codec	459
11.9.1 Introduction	459
11.9.2 The Perceptual Weighting Filter	460
11.9.3 The Open Loop Pitch Search	460
11.9.4 The Closed Loop Pitch Search	460
11.9.5 The Algebraic Codebook Search	461
11.9.6 The Decoder Post Processing	461
11.9.7 Conclusions	462
11.10 The Enhanced Full-rate GSM codec	462
11.10.1 Codec Outline	462
11.10.2 Operation of the EFR-GSM Encoder	464
11.10.2.1 Spectral Quantisation in the EFR-GSM Codec	464
11.10.2.2 Adaptive Codebook Search	466
11.10.2.3 Fixed Codebook Search	467
11.11 The IS-136 Speech Codec	468
11.11.1 IS-136 codec outline	468
11.11.2 IS-136 Bitallocation scheme	469
11.11.3 Fixed Codebook Search	471
11.11.4 IS-136 Channel Coding	472
11.12 The ITU G.723.1 Dual-Rate Codec	473
11.12.1 Introduction	473
11.12.2 G.723.1 Encoding Principle	473
11.12.3 Vector-Quantisation of the LSPs	476
11.12.4 Formant-based Weighting Filter	476
11.12.5 The 6.3 kbps High-rate G.723.1 Excitation	477
11.12.6 The 5.3 kbps low-rate G.723.1 excitation	479
11.12.7 G.723.1 Bitallocation	480
11.12.8 G.723.1 Error Sensitivity	481
11.13 Summary of Standard CELP-based Codecs	483
12 Backward-Adaptive CELP Coding	487
12.1 Introduction	487
12.2 Motivation and Background	488
12.3 Backward-Adaptive G728 Schematic	492
12.4 Backward-Adaptive G728 Coding	493
12.4.1 G728 Error Weighting	493
12.4.2 G728 Windowing	493
12.4.3 Codebook Gain Adaption	497
12.4.4 G728 Codebook Search	500
12.4.5 G728 Excitation Vector Quantization	503
12.4.6 G728 Adaptive Postfiltering	505
12.4.6.1 Adaptive Long-term Postfiltering	505
12.4.6.2 G728 Adaptive Short-term Postfiltering	507
12.4.7 Complexity and Performance of the G728 Codec	508
12.5 Reduced-Rate 16-8 kbps G728-Like Codec I	509

CONTENTS

ix

12.6	The Effects of Long Term Prediction	512
12.7	Closed-Loop Codebook Training	517
12.8	Reduced-Rate 16-8 kbps G728-Like Codec II	522
12.9	Programmable-Rate 8-4 kbps CELP Codecs	524
12.9.1	Motivation	524
12.9.2	8-4kbps Codec Improvements	524
12.9.3	8-4kbps Codecs - Forward Adaption of the STP Synthesis Filter	525
12.9.4	8-4kbps Codecs - Forward Adaption of the LTP	527
12.9.4.1	Initial Experiments	527
12.9.4.2	Quantization of Jointly Optimized Gains	529
12.9.4.3	8-4kbps Codecs - Voiced/Unvoiced Codebooks	532
12.9.5	Low Delay Codecs at 4-8 kbits/s	534
12.9.6	Low Delay ACELP Codec	537
12.10	Backward-adaptive Error Sensitivity Issues	540
12.10.1	The Error Sensitivity of the G728 Codec	540
12.10.2	The Error Sensitivity of Our 4-8 kbits/s Low Delay Codecs	542
12.10.3	The Error Sensitivity of Our Low Delay ACELP Codec	547
12.11A	Low-Delay Multimode Speech Transceiver	547
12.11.1	Background	547
12.11.2	8-16 kbps Codec Performance	548
12.11.3	Transmission Issues	550
12.11.3.1	Higher-quality Mode	550
12.11.3.2	Lower-quality Mode	552
12.11.4	Speech Transceiver Performance	552
12.12	Chapter Conclusions	552

IV Wideband and Sub-4kbps Coding and Transmission 555

13	Wideband Speech Coding	557
13.1	Subband-ADPCM Wideband Coding	557
13.1.1	Introduction and Specifications	557
13.1.2	G722 Codec Outline	558
13.1.3	Principles of Subband Coding	561
13.1.4	Quadrature Mirror Filtering	563
13.1.4.1	Analysis Filtering	563
13.1.4.2	Synthesis Filtering	566
13.1.4.3	Practical QMF Design Constraints	567
13.1.5	G722 Adaptive Quantisation and Prediction	573
13.1.6	G722 Coding Performance	575
13.2	Wideband Transform-Coding at 32 kbps	575
13.2.1	Background	575
13.2.2	Transform-Coding Algorithm	576
13.3	Subband-Split Wideband CELP Codecs	579
13.3.1	Background	579
13.3.2	Subband-based Wideband CELP coding	580

13.3.2.1	Motivation	580
13.3.2.2	Low-band Coding	580
13.3.2.3	Highband Coding	582
13.3.2.4	Bit allocation Scheme	582
13.4	Fullband Wideband ACELP Coding	583
13.4.1	Wideband ACELP Excitation	583
13.4.2	Wideband 32 kbps ACELP Coding	586
13.4.3	Wideband 9.6 kbps ACELP Coding	587
13.5	Turbo-coded Wideband Speech Transceiver	588
13.5.1	Background and Motivation	588
13.5.2	System Overview	591
13.5.3	System Parameters	592
13.5.4	Constant Throughput Adaptive Modulation	593
13.5.5	Adaptive Wideband Transceiver Performance	595
13.5.6	Multi-mode Transceiver Adaptation	597
13.5.7	Transceiver Mode Switching	598
13.5.8	The Wideband PictureTel Codec	599
13.5.8.1	Audio Codec Overview	599
13.5.9	Detailed Description of the Audio Codec	601
13.5.10	Wideband Adaptive System Performance	603
13.5.11	Audio Frame Error Results	603
13.5.12	Audio Segmental SNR Performance and Discussions	604
13.5.13	Picturetel Audio Transceiver Summary and Conclusions	605
13.6	Chapter Conclusions	606
14	Overview of Speech Coding	609
14.1	Low Bit Rate Speech Coding	609
14.1.1	Analysis-by-Synthesis Coding	611
14.1.2	Speech Coding at 2.4kbps	614
14.1.2.1	Background to 2.4kbps Speech Coding	614
14.1.2.2	Frequency Selective Harmonic Coder	615
14.1.2.3	Sinusoidal Transform Coder	617
14.1.2.4	Multiband Excitation Coders	618
14.1.2.5	Subband Linear Prediction Coder	619
14.1.2.6	Mixed Excitation Linear Prediction Coder	620
14.1.2.7	Waveform Interpolation Coder	621
14.1.3	Speech Coding Below 2.4kbps	622
14.2	Linear Predictive Coding model	624
14.2.1	Short Term Prediction	625
14.2.2	Long Term Prediction	627
14.2.3	Final Analysis-by-Synthesis Model	627
14.3	Speech Quality Measurements	627
14.3.1	Objective Speech Quality Measures	628
14.3.2	Subjective Speech Quality Measures	629
14.3.3	2.4kbps Selection Process	629
14.4	Speech Database	631

CONTENTS

xi

14.5 Summary	634
15 Linear Predictive Vocoder	635
15.1 Overview of a Linear Predictive Vocoder	635
15.2 Line Spectrum Frequencies Quantization	636
15.2.1 Line Spectrum Frequencies Scalar Quantization	636
15.2.2 Line Spectrum Frequencies Vector Quantization	638
15.3 Pitch Detection	642
15.3.1 Voiced-Unvoiced Decision	643
15.3.2 Oversampled Pitch Detector	645
15.3.3 Pitch Tracking	647
15.3.3.1 Computational Complexity	651
15.3.4 Integer Pitch Detector	654
15.4 Unvoiced Frames	654
15.5 Voiced Frames	655
15.5.1 Placement of Excitation Pulses	656
15.5.2 Pulse Energy	656
15.6 Adaptive Postfilter	656
15.7 Pulse Dispersion Filter	659
15.7.1 Pulse Dispersion Principles	659
15.7.2 Pitch Independent Glottal Pulse Shaping Filter	661
15.7.3 Pitch Dependent Glottal Pulse Shaping Filter	662
15.8 Results for Linear Predictive Vocoder	664
15.9 Summary and Conclusions	669
16 Wavelets and Pitch Detection	671
16.1 Conceptual Introduction to Wavelets	671
16.1.1 Fourier Theory	671
16.1.2 Wavelet Theory	673
16.1.3 Detecting Discontinuities with Wavelets	673
16.2 Introduction to Wavelet Mathematics	675
16.2.1 Multiresolution Analysis	675
16.2.2 Polynomial Spline Wavelets	677
16.2.3 Pyramidal Algorithm	677
16.2.4 Boundary Effects	679
16.3 Preprocessing the Wavelet Transform Signal	679
16.3.1 Spurious Pulses	680
16.3.2 Normalization	680
16.3.3 Candidate Glottal Pulses	680
16.4 Voiced-Unvoiced Decision	683
16.5 Wavelet Based Pitch Detector	684
16.5.1 Dynamic Programming	685
16.5.2 Autocorrelation Simplification	688
16.6 Summary and Conclusions	692

17 Zinc Function Excitation	693
17.1 Introduction	693
17.2 Overview of Prototype Waveform Interpolation Zinc Function Excitation	695
17.2.1 Coding Scenarios	695
17.2.1.1 U-U-U Encoder Scenario	695
17.2.1.2 U-U-V Encoder Scenario	695
17.2.1.3 V-U-U Encoder Scenario	697
17.2.1.4 U-V-U Encoder Scenario	697
17.2.1.5 V-V-V Encoder Scenario	698
17.2.1.6 V-U-V Encoder Scenario	698
17.2.1.7 U-V-V Encoder Scenario	698
17.2.1.8 V-V-U Encoder Scenario	698
17.2.1.9 U-V Decoder Scenario	698
17.2.1.10 U-U Decoder Scenario	700
17.2.1.11 V-U Decoder Scenario	700
17.2.1.12 V-V Decoder Scenario	700
17.3 Zinc Function Modelling	700
17.3.1 Error Minimization	701
17.3.2 Computational Complexity	702
17.3.3 Reducing the Complexity of Zinc Function Excitation Opti- mization	702
17.3.4 Phases of the Zinc Functions	704
17.4 Pitch Detection	704
17.4.1 Voiced-Unvoiced Boundaries	704
17.4.2 Pitch Prototype Selection	705
17.5 Voiced Speech	708
17.5.1 Energy Scaling	710
17.5.2 Quantization	711
17.6 Excitation Interpolation Between Prototype Segments	713
17.6.1 ZFE Interpolation Regions	713
17.6.2 ZFE Amplitude Parameter Interpolation	714
17.6.3 ZFE Position Parameter Interpolation	714
17.6.4 Implicit Signalling of Prototype Zero Crossing	715
17.6.5 Removal of ZFE Pulse Position Signalling and Interpolation . .	716
17.6.6 Pitch Synchronous Interpolation of Line Spectrum Frequencies	716
17.6.7 ZFE Interpolation Example	717
17.7 Unvoiced Speech	717
17.8 Adaptive Postfilter	717
17.9 Results for Single Zinc Function Excitation	720
17.10 Error Sensitivity of the 1.9kbps PWI-ZFE Coder	723
17.10.1 Parameter Sensitivity of the 1.9kbps PWI-ZFE coder	724
17.10.1.1 Line Spectrum Frequencies	724
17.10.1.2 Voiced-Unvoiced Flag	724
17.10.1.3 Pitch Period	724
17.10.1.4 Excitation Amplitude Parameters	725
17.10.1.5 Root Mean Square Energy Parameter	725

CONTENTS

xiii

17.10.1.6 Boundary Shift Parameter	725
17.10.2 Degradation from Bit Corruption	725
17.10.2.1 Error Sensitivity Classes	727
17.11 Multiple Zinc Function Excitation	727
17.11.1 Encoding Algorithm	728
17.11.2 Performance of Multiple Zinc Function Excitation	730
17.12A Sixth-rate, 3.8 kbps GSM-like Speech Transceiver	734
17.12.1 Motivation	734
17.12.2 The Turbo-coded Sixth-rate 3.8 kbps GSM-like System	735
17.12.3 Turbo Channel Coding	736
17.12.4 The Turbo-coded GMSK Transceiver	738
17.12.5 System Performance Results	739
17.13 Summary and Conclusions	740
18 Mixed-Multiband Excitation	741
18.1 Introduction	741
18.2 Overview of Mixed-Multiband Excitation	742
18.3 Finite Impulse Response Filter	744
18.4 Mixed-Multiband Excitation Encoder	748
18.4.1 Voicing Strengths	749
18.5 Mixed-Multiband Excitation Decoder	753
18.5.1 Adaptive Postfilter	754
18.5.2 Computational Complexity	754
18.6 Performance of the Mixed-Multiband Excitation Coder	756
18.6.1 Performance of a Mixed-Multiband Excitation Linear Predic- tive Coder	757
18.6.2 Performance of a Mixed-Multiband Excitation and Zinc Func- tion Prototype Excitation Coder	762
18.7 A Higher Rate 3.85kbps Mixed-Multiband Excitation Scheme	766
18.8 A 2.35 kbit/s Joint-detection CDMA Speech Transceiver	768
18.8.1 Background	768
18.8.2 The Speech Codec's Bit Allocation	769
18.8.3 The Speech Codec's Error Sensitivity	770
18.8.4 Channel Coding	770
18.8.5 The JD-CDMA Speech System	771
18.8.6 System performance	772
18.8.7 Conclusions on the JD-CDMA Speech Transceiver	774
18.9 Conclusion	774
19 Sinusoidal Transform Coding Below 4kbps	777
19.1 Introduction	777
19.2 Sinusoidal Analysis of Speech Signals	778
19.2.1 Sinusoidal Analysis with Peak Picking	778
19.2.2 Sinusoidal Analysis using Analysis-by-Synthesis	779
19.3 Sinusoidal Synthesis of Speech Signals	780
19.3.1 Frequency, Amplitude and Phase Interpolation	780

19.3.2	Overlap-Add Interpolation	781
19.4	Low Bit Rate Sinusoidal Coders	784
19.4.1	Increased Frame Length	784
19.4.2	Incorporating Linear Prediction Analysis	785
19.5	Incorporating Prototype Waveform Interpolation	786
19.6	Encoding the Sinusoidal Frequency Component	787
19.7	Determining the Excitation Components	790
19.7.1	Peak-Picking of the Residual Spectra	790
19.7.2	Analysis-by-Synthesis of the Residual Spectrum	790
19.7.3	Computational Complexity	792
19.7.4	Reducing the Computational Complexity	793
19.8	Quantizing the Excitation Parameters	797
19.8.1	Encoding the Sinusoidal Amplitudes	797
19.8.1.1	Vector Quantization of the Amplitudes	797
19.8.1.2	Interpolation and Decimation	797
19.8.1.3	Vector Quantization	800
19.8.1.4	Vector Quantization Performance	800
19.8.1.5	Scalar Quantization of the Amplitudes	801
19.8.2	Encoding the Sinusoidal Phases	803
19.8.2.1	Vector Quantization of the Phases	803
19.8.2.2	Encoding the Phases with a Voiced-Unvoiced Switch	803
19.8.3	Encoding the Sinusoidal Fourier Coefficients	804
19.8.3.1	Equivalent Rectangular Bandwidth Scale	804
19.8.4	Voiced-Unvoiced Flag	806
19.9	Sinusoidal Transform Decoder	806
19.9.1	Pitch Synchronous Interpolation	807
19.9.1.1	Fourier Coefficient Interpolation	807
19.9.2	Frequency Interpolation	807
19.9.3	Computational Complexity	807
19.10	Speech Coder Performance	808
19.11	Summary and Conclusions	814
20	Conclusions on Low Rate Coding	817
20.1	Summary	817
20.2	Listening Tests	818
20.3	Conclusions	820
20.4	Further Research	821
21	Comparison of Speech Transceivers	823
21.1	Background to Speech Quality Evaluation	823
21.2	Objective Speech Quality Measures	824
21.2.1	Introduction	824
21.2.2	Signal to Noise Ratios	825
21.2.3	Articulation Index	826
21.2.4	Cepstral Distance	826
21.2.5	Cepstral Example	829

CONTENTS	1
21.2.6 Logarithmic likelihood ratio	831
21.2.7 Euclidean Distance	832
21.3 Subjective Measures	832
21.3.1 Quality Tests	833
21.4 Comparison of Quality Measures	834
21.4.1 Background	834
21.4.2 Intelligibility tests	835
21.5 Subjective Speech Quality of Various Codecs	836
21.6 Speech Codec Bit-sensitivity	837
21.7 Transceiver Speech Performance	844
A Constructing the Quadratic Spline Wavelets	847
B Zinc Function Excitation	851
C Probability Density Function for Amplitudes	855
Bibliography	861
Index	895
Author Index	895

Preface and Motivation

The Speech Coding Scene

In the era of the third - generation (3G) wireless personal communications standards - despite the emergence of broad-band access network standard proposals - the most important mobile radio services are still based on voice communications. 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 - predominantly after compression.

This book is dedicated mainly to 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 incorporated in source-sensitivity matched wireless transceivers. 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 implementational 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.

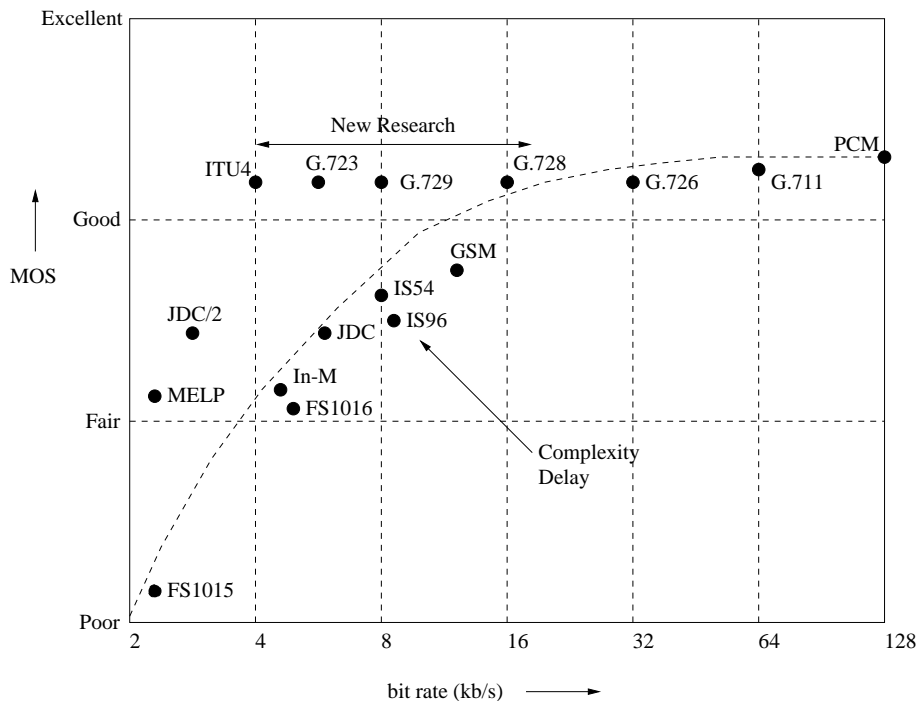


Figure 1: Subjective speech quality of various codecs [1] ©IEEE, 1996

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. Work is under way at the time of writing towards the standardisation of a 4 kbps ITU scheme, which we refer to here as ITU4.

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-Synchronous 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 became known in the community, which led to the requirement of designing an appropriate multi-mode codec, the Advanced Multi-Rate codec referred to as the AMR codec. A range of IMTs also constitute the subject of this book. Current research on sub-2.4 kbps speech codecs is also covered extensively in the book, where the aspects of auditory masking become more dominant. Lastly, since the classic G.722 subband-ADPCM based wideband codec is becoming somewhat obsolete in the light of exciting new development in compression, the most recent trend is to consider wideband speech and audio codecs, providing substantially 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 PictureTel codec - which can be programmed to operate between 10 kbps and 32 kbps and hence amenable to employment in IMTs - 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 good 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 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 laid 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], Pappamichalis [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].

During the early nineties Atal, Cuperman and Gersho [28] have edited prestigious contributions on speech compression. Also Ince [29] contributed a book in 1992 related to the topic. Anderson and Mohan co-authored a monograph on source and channel coding in 1993 [30]. Most of the recent developments were then consolidated in Kondoz' excellent monograph in 1994 [31] and in the multi-authored contribution edited by Keijn and Paliwal [32] in 1995. The most recent addition to the above range of contributions is the second edition of O'Shaughnessy well-referenced book cited above.

Motivation and Outline of the Book

Against this backcloth - since the publication of Kondoz's monograph in 1994 [31] nearly six years have elapsed - this book endeavours to review the recent history of speech compression and communications. 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

CONTENTS

7

various speech codecs studied also in the context of a full wireless transceiver.

The book is constituted by four parts. Part I and II are covering classic background material, while the bulk of the book is constituted by the research-oriented Part III and IV, covering both standardised and proprietary speech codecs and transceivers. Specifically, Part I provides a rudimentary introduction to the wireless system components used throughout the book in quantifying the overall performance of the various speech codecs, in order to render our treatment of the topics self-contained. Specifically, the mobile propagation environment, modulation and transmission techniques as well as channel coding are considered in Chapters 1-4. For the sake of completeness Part II focusses on aspects of classic waveform coding and predictive coding in Chapters 5 and 6. Part III is centred around analysis by synthesis based coding, reviewing the principles in Chapter 7 as well as both narrow and wideband spectral quantisation in Chapter 8. RPE and CELP coding are the topic of Chapters 9 and 10, which are followed by an approximately 100-page chapter on the existing forward-adaptive standard CELP codecs in Chapter 11 and on their associated source-sensitivity matched channel coding schemes. The subject of Chapter 12 is 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 essentially research-oriented Part IV is dedicated to a range of standard and proprietary wideband, as well as sub-4kbps coding techniques and wireless systems. As an introduction to the scene, the classic G.722 wideband codec is reviewed first, leading to various low-rate wideband codecs. Chapter 13 is concluded with a turbo-coded Orthogonal Frequency Division Multiplex (OFDM) wideband audio system design example. The remaining chapters, namely Chapters 14-21 are all dedicated to sub-4kbps codecs and transceivers.

This book is naturally limited in terms of its coverage of these aspects, simply due 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.

We hope that the book offers you 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 unbiased 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 with intellectual stimulation.

Lajos Hanzo

Acknowledgements

The book has been written by the staff in the Electronics and Computer Science Department at the University of Southampton. We are indebted to our many colleagues who have enhanced our understanding of the subject. 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 enlightenment gained from our collaborations on various projects, papers and books. We are grateful to J. Brecht, Jon Blogh, Marco Breiling, M. del Buono, Clare Brooks, Stanley Chia, Byoung Jo Choi, Joseph Cheung, Peter Fortune, Lim Dongmin, D. Didascalou, S. Ernst, Eddie Green, David Greenwood, Hee Thong How, Thomas Keller, W.H. Lam, C.C. Lee, M.A. Nofal, Xiao Lin, Chee Siong Lee, Tong-Hooi Liew, Matthias Muenster, V. Roger-Marchart, Redwan Salami, David Stewart, Jeff Torrance, Spiros Vlahoyiannatos, 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.

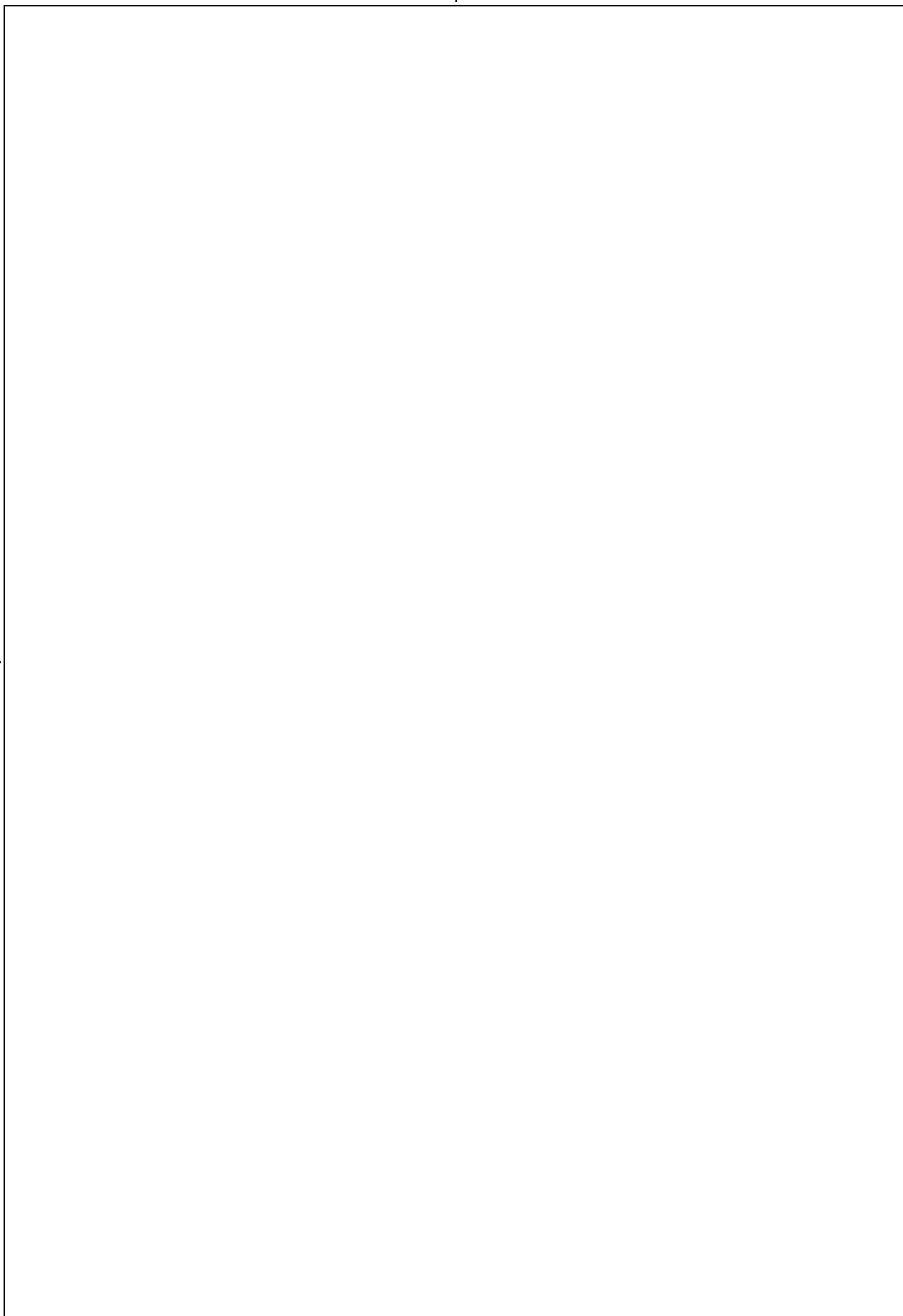
We also acknowledge our valuable associations with the Virtual Centre of Excellence in Mobile Communications, in particular with its Chief Executives, Dr. Tony Warwick and Dr Walter Tuttlebee as well as with Dr. Keith Baughan and other members of its Executive Committee. Our sincere thanks are also due to the EPSRC, UK; Dr. Joao Da Silva, Dr Jorge Pereira and other colleagues from the Commission of the European Communities, Brussels; Andy Wilton, Luis Lopes and Paul Crichton from Motorola ECID, Swindon, UK for sponsoring some of our recent research.

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

Part I

Transmission Issues



162

486

Chapter 12

Backward-Adaptive Code Excited Linear Prediction

12.1 Introduction

In the previous chapter a range of medium-to-high delay forward-adaptive CELP codecs have been described, which constituted different trade-offs in terms of speech quality, bitrate, delay and implementational complexity. In this chapter our work moves on to low delay, backward adaptive, codecs.

The outline of this chapter is as follows. In the next Section we discuss why the delay of a speech codec is an important parameter, methods of achieving low delay coding and problems with these methods. Much of the material presented is centered around the recently standardised 16 kbits/s G728 Low Delay CELP codec [217, 231], and the associated algorithmic issues are described in Section 12.4. We then describe our attempts to extend the G728 codec in order to propose a low delay, programmable bit rate codec operating between 8 kbits/s and 16 kbits/s. In Section 12.6 we describe the potential speech quality improvements that can be achieved in such a codec by adding a Long Term Predictor (LTP), albeit at the cost of increased error sensitivity due to error propagation effects introduced by the backward-adaptive LTP. These error propagation effects can be mitigated at system level, for example by introducing reliable error control mechanisms, such as Automatic Repeat Request (ARQ), an issue to be discussed in a system context at a later stage. In Section 12.7 we discuss means of training the codebooks used in our variable rate codec to optimise its performance. Section 12.8 describes an alternative variable rate codec which has a constant vector size. Finally in Section 12.4.6 we describe the postfiltering which is used to improve the perceptual quality of our codecs.

12.2 Motivation and Background

The delay of a speech codec can be an important parameter for several reasons. In the public switched telephone network 4 to 2 wire conversions lead to echoes, which will be subjectively annoying if the echo is sufficiently delayed. Experience shows that the 57.5 ms speech coding and interleaving delay of the Pan-European GSM system already introduces an undesirable echoing effect and this value can be considered as the maximum tolerable margin in toll-quality communications. Even if echo cancellers are used, a high delay speech codec makes the echo cancellation more difficult. Therefore if a codec is to be connected to the telephone network it is desirable that its delay should be as low as possible. If the speech codec used has a lower delay, then other elements of the system, such as bit inter-leavers, will have more flexibility and should be able to improve the overall quality of the system.

The one-way **coding delay** of a speech codec is defined as the time from when a sample arrives at the input of the encoder to when the corresponding sample is produced at the output of the decoder, assuming the bit-stream from the encoder is fed directly to the decoder. This one-way delay is typically made up of three main components [217]. The first is the **algorithmic buffering delay** of the codec - the encoder operates on frames of speech, and must buffer a frame-lengths worth of speech samples before it can start encoding. The second component of the overall delay is the **processing delay** - speech codecs typically operate in just real time, and so it takes almost one frame length in time to process the buffered samples. Finally there is the bit **transmission delay** - if the encoder is linked to the decoder by a channel with capacity equal to the bit rate of the codec then there will be a further time delay equal to the codec's frame length while the decoder waits to receive all the bits representing the current frame.

From the above description the overall one-way delay of the codec will be equal to about three times the frame length of the codec. However it is possible to reduce this delay by careful implementation of the codec. For example if a faster processor is used the processing delay can be reduced. Also it may not be necessary to wait until the whole speech frame has been processed before we can start sending bits to the decoder. Finally a faster communications channel, for example in a time division multiplexed system, can dramatically reduce the bit transmission delay. Other factors may also result in the total delay being increased. For example the one sub-frame look-ahead used to aid the interpolation of the LSFs in our ACELP codecs described earlier will increase the overall delay by one sub-frame. Nonetheless, typically the one-way coding delay of a speech codec is assumed to be about 2.5 to 3 times the frame length of the codec.

It is obvious from the discussion above that the most effective way of producing a low delay speech codec is to use as short a frame length as possible. Traditional CELP codecs have a frame length of 20 to 30 ms, leading to a total coding delay of at least 50 ms. Such a long frame length is necessary because of the forward adaption of the short-term synthesis filter coefficients. As explained in Chapter 10 a frame of speech is buffered, LPC analysis is performed and the resulting filter coefficients are quantized and transmitted to the decoder. As we reduce the frame length, the filter coefficients must be sent more often to the decoder and so more and more of the available bit

rate is taken up by LPC information. Although efficient speech windowing and LSF quantization schemes have allowed the frame length to be reduced to 10 ms (with a 5 ms look-ahead) in a candidate codec [143] for the CCITT 8 kbits/s standard, a frame length of between 20 and 30 ms is more typical. If we want to produce a codec with delay of the order of 2 ms, which was the objective for the CCITT 16 kbits/s codec [231], it is obvious that we cannot use forward adaption of the synthesis filter coefficients.

The alternative is to use backward adaptive LPC analysis. This means that rather than window and analyse present and future speech samples in order to derive the filter coefficients, we analyse previous quantized and locally decoded signals to derive the coefficients. These past quantized signals are available at both the encoder and decoder, and so no side information about the LPC coefficients needs to be transmitted. This allows us to update the filter coefficients as frequently as we like, with the only penalty being a possible increase in the complexity of the codec. Thus we can dramatically reduce the codec's frame length and delay.

As explained above backward adaptive LPC analysis has the advantages of allowing us to dramatically reduce the delay of our codec, and removing the information about the filter coefficients that must be transmitted. This side information is usually about 25 % of the bit rate of a codec, and so it is very helpful if it can be removed. However backward adaption has the disadvantage that it produces filter coefficients which are typically degraded in comparison to those used in forward adaptive codecs. The degradation in the coefficients comes from two sources [339]:

1. **Noise Feedback** - In a backward adaptive system the filter coefficients are derived from a quantized signal, and so there will be a feedback of quantization noise into the LPC analysis which will degrade the performance of the coefficients produced.
2. **Time Mismatch** - In a forward adaptive system the filter coefficients for the current frame are derived from the input speech signal for the current frame. In a backward adaptive system we have only signals available from previous frames to use, and so there is a time mismatch between the current frame and the coefficients we use for that frame.

The effects of noise feedback especially increases dramatically as the bit rate of the codec is reduced and means that traditionally backward adaption has only been used in high bit rate, high quality, codecs. However recently, as researchers have attempted to reduce the delay of speech codecs, backward adaptive LPC analysis has been used at bit rates as low as 4.8 kbits/s [340].

Clearly, the major design challenge associated with the ITU G728 codec was due to the complexity of its specifications, which are summarised in Table 12.1. Although many speech codecs can produce good speech quality at 16 kbps, at such a low rate most previous codecs have inflicted significantly higher delays than the targeted 2 ms. This is due to the fact that in order to achieve such a low rate in linear predictive coding, the up-date interval of the LPC coefficients must be around 20–30 ms. We have argued before in Section 8 that in case of scalar LPC parameter coding typically 36 bits/20 ms=1.8 kbps channel capacity is required for their encoding. Hence, in

Parameter	Specification
Bitrate	16 kbps
One-way delay	< 2 ms
Speech quality at $BER = 0$	< 4 QDU for one codec < 14 QDU for three tandems
Speech quality at $BER = 10^{-3}$ and 10^{-2}	Better than that of G721 32 kbps ADPCM
Additional requirement	Pass DTMF and CCITT No. 5, 6 and 7 signalling

Table 12.1: G 728 codec specifications

case of a 2 ms delay forward predictive coding is not a realistic alternative. We have also seen in Section 6.9 that low-complexity, low-delay ADPCM coding at 16 kbps is possible, which would satisfy the first two criteria of Table 12.1, but the last three requirements are not satisfied.

Chen, Cox, Lin, Jayant and Melchner have contributed a major development to the state-of-art of speech coding [217], which satisfied all the design specifications and was standardised by the ITU [231]. In this Section we will follow their discussions in References [217] and [231], pp 625-627, in order to describe the operation of their proposed backward adaptive codec. The ITU's call for proposals stimulated a great deal of research, and a variety of candidate codecs were proposed, which typically satisfied some but not all requirements of Table 12.1. Nonetheless, a range of endeavours - amongst others those of References [228, 341] - have contributed in various ways towards the standardisation process.

CELP coding emerged as the best candidate, which relied on backward prediction using a high-order (50) filter, where the coefficients did not have to be transmitted, they were extracted from the past decoded speech. Due to the high-order short-term predictor (STP) there was no need to include an error-sensitive long-term predictor (LTP). The importance of adaptive post-filtering was underlined by Jayant and Ramamoorthy in [228, 229], where the quality of 16 kbps ADPCM-coded speech was reportedly improved, which was confirmed by Chen and Gersho [230].

The delay and high speech quality criteria were achieved by using a short STP-update interval of 20 samples or $20 \cdot 125 \mu s = 2.5 ms$ and an excitation vector length of 5 samples or $5 \cdot 125 \mu s = 0.625 ms$. The speech quality was improved using a trained codebook rather than a stochastic one, which was 'virtually' extended by a factor of eight using a 3-bit codebook gain factor. Lastly, a further novel element of the codec is the employment of backward adaptive gain scaling [342, 343], which will be discussed in more depth during our further discourse. In the next Section we will describe the 16 kbits/s G728 low delay CELP codec, and in particular the ways it differs from the ACELP codecs we have used previously. We will also attempt to quantify the effects of both noise feedback and time mismatch on the backward adaptive LPC analysis used in this codec. Let us now focus our attention on specific details of the codec.

12.2. MOTIVATION AND BACKGROUND

491

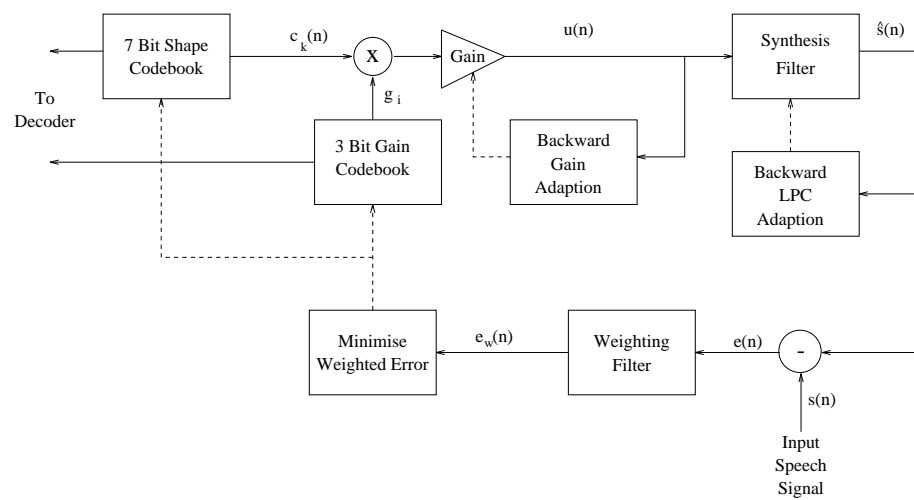


Figure 12.1: 16 kbps low-delay CCITT G728 Encoder

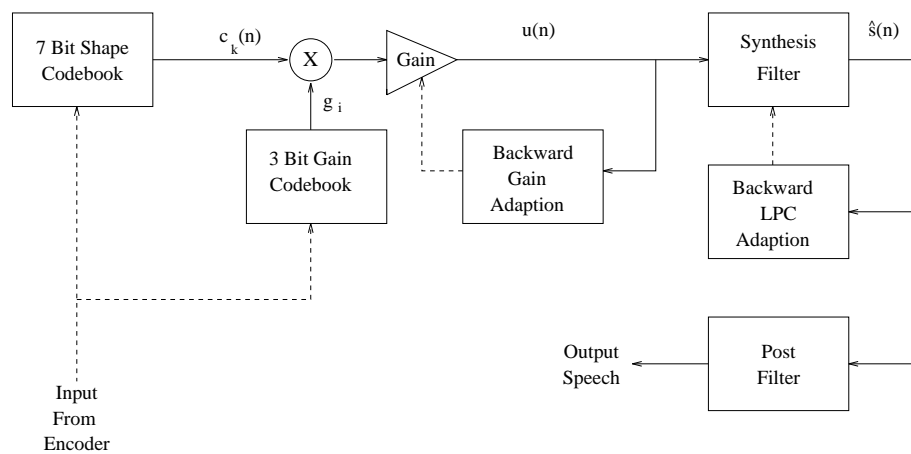


Figure 12.2: 16 kbps low-delay CCITT G728 Decoder

12.3 Backward-Adaptive G728 Codec Schematic [217, 231]

The G728 encoder and decoder schematics are portrayed in Figures 12.1 and 12.2, respectively. The input speech segments are compared with the synthetic speech segments as in any ABS-codec, and the error signal is perceptually weighted, before the specific codebook entry associated with the lowest error is found in an exhaustive search procedure. For the G728 codec a vector size of 5 samples corresponding to $5 \cdot 125\mu s = 0.625ms$ was found appropriate in order to curtail the overall speech delay to 2 ms.

Having fixed the length of the excitation vectors, let us now consider the size of the excitation codebook. Clearly, the larger the codebook size, the better the speech quality, but the higher the computational complexity, and the bitrate. An inherent advantage of backward adaptive prediction is that the LPC coefficients are not transmitted, hence a high-order filter can be used and we can dispense with using an LTP. Therefore, a design alternative is to allocate all bits transmitted to the codebook indices. Assuming a transmission rate of 16 kbps and an 8 kHz sampling rate, we are limited to a coding rate of 2 bits/sample or 10 bits/5 samples. Logically, the maximum possible codebook size is then $2^{10} = 1024$ entries. Recall that in case of forward predictive codecs the codebook gain was typically quantised using 4–5 bits, which allowed a degree of flexibility in terms of excitation envelope fluctuation. In this codec it is unacceptable to dedicate such a high proportion of the bit-rate budget to the gain quantisation. Chen and Gersho [343] noted that this slowly fluctuating gain information is implicitly available and hence predictable on the basis of previously scaled excitation segments. This prompted them to contrive a **backward adaptive gain predictor**, which infers the required current scaling factor from its past values using predictive techniques. The actual design of this gain predictor will be highlighted at a later stage. Suffice to say here that this allowed the total of 10 bits to be allocated to the codebook index, although the codebook finally was trained as a 128-entry scheme in order to reduce the search complexity by a factor of eight, and the remaining three bits were allocated to quantise another multiplicative gain factor. This two-stage approach is suboptimum in terms of coding performance, since it replaces eight independent codebook vectors by eight identically shaped, different magnitude excitation vectors. Nonetheless, the advantage of the eight-fold reduced complexity outweighed the significance of a slight speech degradation.

As mentioned before, Chen *et al.* decided to opt for a 50th order backward adaptive STP filter in order to achieve the highest possible prediction gain, and to be able to dispense with LTP filtering, without having to transmit any LPC coefficients. However, the complexity of the Levinson–Durbin algorithm used to compute the LPC coefficients is proportional to the square of the filter order $p = 50$, which constitutes a high complexity. This is particularly so if the LPC coefficients are updated for each 5-sample speech vector. In order to compromise, an update interval of 20 samples or 2.5 ms was deemed to be appropriate. This implies that the LPC parameters are kept constant for the duration of four excitation vectors, which is justifiable since the speech spectral envelope does not vary erratically.

A further ramification of extending the LPC update interval is that the time-lag

between the speech segment to be encoded and the spectral envelope estimation is increased. This is a disadvantage of backward adaptive predictive systems, since in backward adaptive schemes the current speech frame is used for the speech spectral estimation. On the same note, backward adaptive arrangements have to infer the LPC coefficients from the past decoded speech, which is prone to quantisation effects. In case of high-rate, high-quality coding this is not a significant problem, but it is aggravated by error propagation effects, inflicting future impairments in future LPC coefficients. Hence, at low bitrates, below 8 kbps, backward adaptive schemes found only limited favour in the past. These effects can be readily quantified using the unquantised original delayed speech signal and the quantised but not delayed speech signal to evaluate the codec's performance. Woodard [299] found that the above factors degraded the codec's SEGSNR performance by about 0.2 dB due to quantisation noise feedback, and by about 0.7 dB due to the time mismatch, yielding a total of 0.9 dB SEGSNR degradation. At lower rates and higher delays these degradations become more dominant. Let us now concentrate our attention on specific algorithmic issues of the codec schematics given in Figures 12.1 and 12.2.

12.4 Backward-Adaptive G728 Coding Algorithm [217, 231]

12.4.1 G728 Error Weighting

In contrast to the more conventional **error weighting filter** introduced in Equation 7.8, the G728 codec employs the filter [230]:

$$W(z) = \frac{1 - A(z/\gamma_1)}{1 - A(z/\gamma_2)} = \frac{1 - \sum_{i=1}^{10} a_i \gamma_1^i}{1 - \sum_{i=1}^{10} a_i \gamma_2^i} \quad (12.1)$$

where $\gamma_1 = 0.9$ and $\gamma_2 = 0.6$, and the filter is based on a 10th order LPC analysis carried out using the unquantised input speech. This was necessary to prevent the introduction of spectral distortions due to quantisation noise. Since the error weighting filter is only used at the encoder, where the original speech signal is available, this error weighting procedure does not constitute any problem at all. The choice of the $\gamma_1 = 0.9$ and $\gamma_2 = 0.6$ parameters was motivated by the requirement of optimising the tandemised performance for three asynchronous coding operations. Explicitly, listening tests proved that the pair $\gamma_1 = 0.9$ and $\gamma_2 = 0.4$ gave a better single-coding performance, but for three tandemed codec $\gamma_2 = 0.6$ was found to exhibit a superior performance. The coefficients of this weighting filter are computed from the windowed input speech, and the particular choice of the window function will be highlighted in the next Section.

12.4.2 G728 Windowing

The choice of the windowing function plays an important role in capturing the time-variant statistics of the input speech, which in turn influences the subsequent spectral

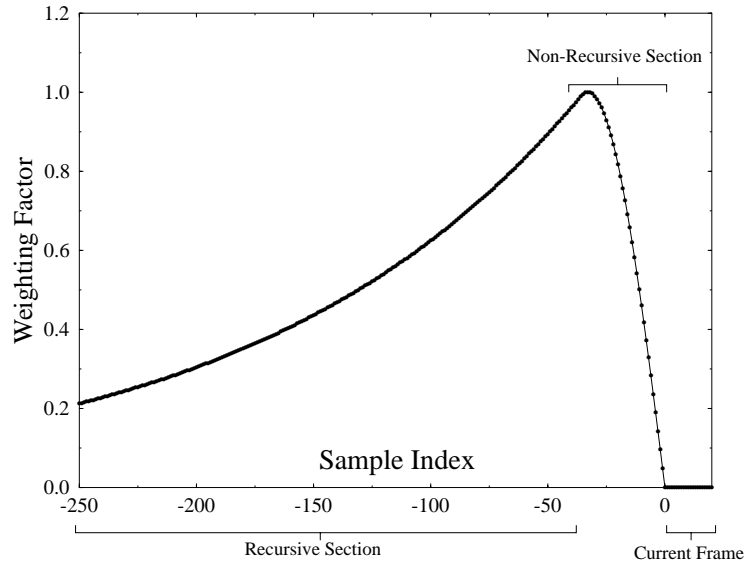


Figure 12.3: Windowing Function Used in the Backward Adaption of the Synthesis Filter

analysis. In contrast to more conventional Hamming windowing, Chen *et al.* [217] proposed to use a hybrid window, which is constituted by an exponentially decaying long-term past history section and a non-recursive section, as it is depicted in Figure 12.3.

Let us assume that the LPC analysis frame size is $L = 20$ samples, which hosts the samples $s(m)$, $s(m+1) \dots s(m+L-1)$, as portrayed in Figure 12.3. The N -sample window section immediately preceding the current LPC frame of L samples is then termed as the non-recursive portion, since it is described mathematically by the help of a sinusoid non-recursive function of $w(n) = -\sin[c(n-m)]$, where the sample index n is limited to the previous N samples ($m-N \leq n \leq (m-1)$). In contrast, the recursive section of the window function weights the input speech samples preceding ($m-N$), as suggested by Figure 12.3, using a simple negative exponential function given by

$$w(n) = b \cdot \alpha^{-[n-(m-N-1)]} \text{ if } n \leq (m-N-1), \quad (12.2)$$

where $0 < b, \alpha < 1$. Evaluating Equation 12.2 for sample index values at the left of $n = (m-N)$ in Figure 12.3 yields weighting factors of $b, b \cdot \alpha, b \cdot \alpha^2 \dots$. In summary, the hybrid window function can be written as:

$$w_m(n) = \begin{cases} f_m(n) = b \cdot \alpha^{-[n-(m-N-1)]} & \text{if } n \leq (m-N-1) \\ g_m(n) = -\sin[c(n-m)] & \text{if } (m-N) \leq n \leq (m-1) \\ 0 & \text{if } n \geq m \end{cases} \quad (12.3)$$

It is important to maintain a seamless transition between the recursive and non-recursive section of the window function in order to avoid introducing spectral side-lobes, which would be incurred in case of a non-continuous derivative at $n = (m - N)$ [344], where the two sections are joined.

Cheng *et al.* also specify in the Recommendation [231] how this recursive windowing process can be exploited to calculate the required autocorrelation coefficients, using the windowed speech signal given by:

$$s_m(n) = s(n) \cdot w_m(n) \quad (12.4)$$

where the subscript m indicates the commencement of the current L -sample window in Figure 12.3.

In case of an M th order LPC analysis at instant m , the autocorrelation coefficients $R_m(i)$ $i = 0, 1, 2 \dots M$ are required by the Levinson–Durbin algorithm, where

$$\begin{aligned} R_m(i) &= \sum_{n=-\infty}^{m-1} s_m(n) \cdot s_m(n-i) \\ &= \sum_{n=-\infty}^{m-N-1} s_m(n) \cdot s_m(n-i) + \sum_{n=m-N}^{m-1} s_m(n) \cdot s_m(n-i). \end{aligned} \quad (12.5)$$

Upon taking into account Equations 12.3 and 12.4 in Equation 12.5, the first term of Equation 12.5 can be written as follows:

$$r_m(i) = \sum_{n=-\infty}^{m-N-1} s(n) \cdot s(n-i) \cdot f_m(n) \cdot f_m(n-i), \quad (12.6)$$

which constitutes the recursive component of $R_m(i)$, since it is computed from the recursively weighted speech segment. The second term of Equation 12.5 relates to the section given by $(m - N) \leq n \leq (m - 1)$ in Figure 12.3, which is the non-recursive section. The N -component sum of the second term is computed for each new N -sample speech segment, while the recursive component can be calculated recursively following the procedure proposed by Chen *et al* [217, 231] as outlined below.

Assuming that $r_m(i)$ is known for the current frame we proceed to the frame commencing at sample position $(m + L)$, which corresponds to the next frame in Figure 12.3, and express $r_{m+L}(i)$ in analogy with Equation 12.5 as follows:

$$\begin{aligned} r_{m+L}(i) &= \sum_{n=-\infty}^{m-1} s_m(n) \cdot s_m(n-i) \\ &= \sum_{n=-\infty}^{m-N-1} s_m(n) \cdot s_m(n-i) + \sum_{n=m-N}^{m-1} s_m(n) \cdot s_m(n-i). \end{aligned}$$

Filter Order p	Δ Prediction Gain (dB)	Δ Seg-SNR (dB)
10	0.0	0.0
25	+0.68	+0.70
50	+1.05	+1.21
75	+1.12	+1.41
100	+1.11	+1.46
150	+1.10	+1.42

Table 12.2: Relative Performance of the Synthesis Filter as p is Increased

$$\begin{aligned}
&= \sum_{n=-\infty}^{m-N-1} s(n) \cdot f_m(n) \cdot \alpha^L \cdot s(n-i) f_m(n-i) \alpha^L \\
&\quad + \sum_{n=m-N}^{m+L-N-1} s_{m+L}(n) \cdot s_{m+L}(n-i) \\
&= L^{2L} r_m(i) + \sum_{n=-N}^{m+L-N-1} s_{m+L}(n) \cdot s_{m+L}(n-i) \quad (12.7)
\end{aligned}$$

This expression is the required recursion, which facilitates the computation of $r_{m+L}(i)$ on the basis of $r_m(i)$. Finally, the total autocorrelation coefficient $R_{m+L}(i)$ is generated by the help of Equation 12.5. When applying the above general hybrid windowing process to the LPC analysis associated with the error weighting, the following parameters are used: $M = 10$, $L = 20$, $N = 30$, $\alpha = (1/2)^{40} \approx 0.983$, yielding $\alpha^{2L} = \alpha^{40} = 1/2$. Then the Levinson and Durbin algorithm is invoked in the usual manner, as described by Equation 6.24 and by the flow chart of Figure 6.3.

The performance of the synthesis filter, in terms of its prediction gain and the segmental SNR of the G728 codec using this filter, is shown against the filter order p in Figure 12.4 for a single sentence spoken by a female. Also shown in Table 12.2 is the increase in performance obtained when p is increased above 10, which is the value most commonly used in AbS codecs. It can be seen that there is a significant performance gain due to increasing the order from 10 to 50, but little additional gain is achieved as p is further increased.

We also tested the degradations in the synthesis filter's performance at $p = 50$ due to backward adaption being used. This was done as follows. To measure the effect of quantization noise feedback we updated the synthesis filter parameters exactly as in G728 except we used the previous speech samples rather than the previous reconstructed speech samples. To measure the overall effect of backward adaption we updated the synthesis filter using both past and present speech samples. The improvements obtained, in terms of the segmental SNR of the codec and the filter's prediction gain, are shown in Table 12.3. We see that due to the high SNR of the G728 codec noise feedback has relatively little effect on the performance of the synthesis filter. The time-mismatch gives a more significant degradation in the codec's performance. Note however that the forward adaptive figures given in Table 12.3 could not be obtained in reality because they do not include any effects of the LPC

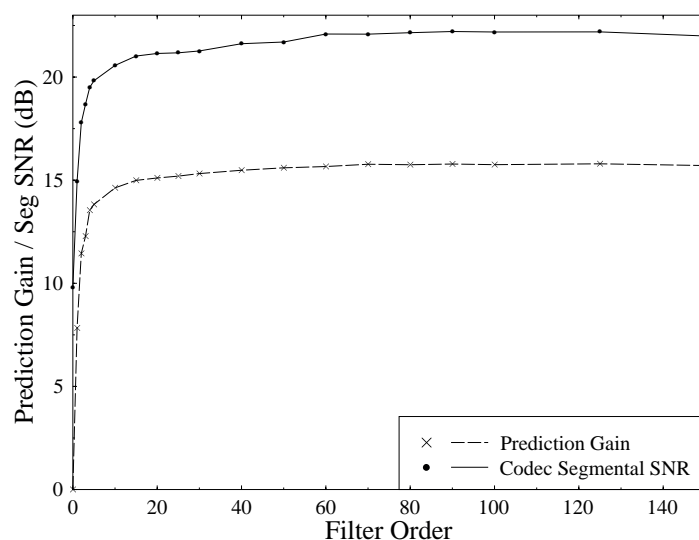


Figure 12.4: Performance of the Synthesis Filter in a G728-Like Codec

	Δ Prediction Gain (dB)	Δ Seg-SNR (dB)
No Noise Feedback	+0.50	+0.18
No Time Mismatch	+0.74	+0.73
Use Forward Adaption	+1.24	+0.91

Table 12.3: Effects of Backward Adaption of the Synthesis Filter

quantization that must be used in a real forward adaptive system.

Having familiarised ourselves with the hybrid windowing process in general terms we note that this process is invoked during three different stages of the G728 codec's operation. The next scheme, where it is employed using a different set of parameters is the code book gain adaption arrangement, which will be elaborated on in the forthcoming Section.

12.4.3 Codebook Gain Adaption

Let us describe the codebook vector scaling process at iteration n by the help of:

$$e(n) = \delta(n) \cdot y(n), \quad (12.8)$$

where $y(n)$ represents one of the 1029 5-sample codebook vectors, $\delta(n)$ the scaling gain factor and $l(n)$ the scaled excitation vector. The associated root-mean-squared

(RMS) values are denoted by $\delta_e(n)$ and $\delta_y(n)$, respectively. As regards to the RMS values we also have:

$$\delta_e(n) = \delta(n) \cdot \delta_y(n) \quad (12.9)$$

or in logarithmic domain:

$$\log[\delta_e(n)] = \log[\delta(n)] + \log[\delta_y(n)].$$

The philosophy of the gain prediction scheme is to exploit the correlation between the current required value of $\delta(u)$ and its past history, which is a consequence of the slowly varying speech envelope. Chen and his colleagues suggested to employ a 10-th order predictor operatin on the sequence $\log[\delta_e(n-1)]$, $\log[\delta_e(n-2)]$... $\log[\delta_e(n-10)]$ in order to predict $\log[\delta(n)]$. This can be written more formally as:

$$\log[\delta(n)] = \sum_{i=1}^{10} p_i \log[\delta_e(n-i)], \quad (12.10)$$

where the coefficient p_i $i = 1...10$ are the predictor coefficients.

When using a 10-th order predictor relying on 10 gain estimates derived for 5 speech samples each the memory of this scheme is 50 samples, which is identical to that of the STP. This predictor therefore analyses the same time interval as the STP and assists in modelling any latent residual pitch periodicity. The excitation gain is predicted for each speech vector n from the 10 previous gain values on the basis of the current set of predictor coefficients p_i $i = 1...10$. These coefficients are then updated using conventional LPC analysis every fourth 5-sample speech vector, or every 20 samples.

The schematic of the gain prediction scheme is depicted in Figure 12.5, where the gain-scaled excitation vector $e(n)$ is buffered and the logarithm of its RMS value is computed in order to express it in terms of dB. At this stage the average excitation gain of voiced speech, namely an offset of 32dB is subtracted in order to remove the bias of the process, before hybrid windowing and LPC analysis takes place.

The **bandwidth expansion** module modifies the predictor coefficients $\hat{\alpha}_i$ computed according to

$$\alpha_i = \left(\frac{29}{32}\right)^i \hat{\alpha}_i = (0.90625)^i \hat{\alpha}_i \quad i = 1...10. \quad (12.11)$$

It can be shown that this process is equivalent in z-domain to moving all the poles of the corresponding synthesis filter towards the origin according to the factor $\left(\frac{29}{32}\right)$. Poles outside the unit circle imply instability, while those inside but close to the unit circle are associated with narrow but high spectral prominances. Moving these poles further away from the unit circle expands their bandwidth and mitigates the associated spectral peaks. If the encoder and decoder are misaligned, for example because the decoder selected te wrong codebook vector due to channel errors, both the speech synthesis filter and the gain prediction scheme will be 'deceived'. The above bandwidth expansion process assists in reducing the error sensitivity of the predictive coefficients by artificially modifying them at both the encoder and decoder using a near-unity leakage factor.

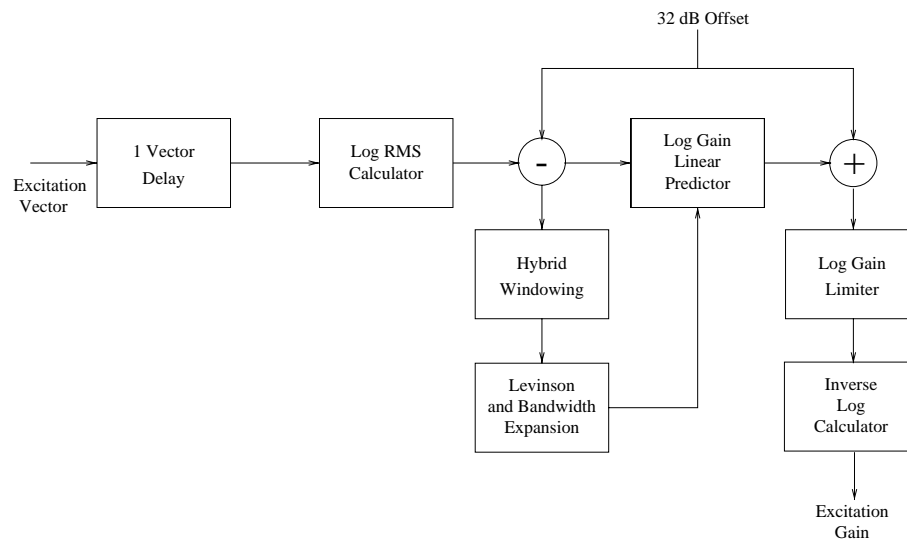


Figure 12.5: G728 Excitation Gain Predictor Scheme

Returning to Figure 12.5, finally the modified predictor coefficients of Equation 12.11 are employed to predict the required logarithmic gain $\log[\sigma(n)]$. Before the gain factor is used in the current frame, its 32 dB offset must be restored, while its extreme values are limited to the range of 0-60 dB and finally $\sigma(n)$ is restored from the logarithmic domain. The linear gain factor is limited accordingly to the range 1-1000.

The efficiency of the backward gain adaption can be seen from Figure 12.6 . This shows the PDFs, on a log scale for clarity, of the excitation vector's optimum gain both with and without gain adaption. Here the optimum vector gain is defined as

$$\sqrt{\frac{1}{vs} \sum_{n=0}^{vs} g^2 c_k^2(n)} \quad (12.12)$$

where g is the unquantized gain chosen in the codebook search. For a fair comparison both PDFs were normalised to have a mean of one. It can be seen that gain adaption produces a PDF which peaks around one and has a shorter tail and a reduced variance. This makes the quantization of the excitation vectors significantly easier. Shown in Figure 12.7 are the PDFs of the optimum unquantized codebook gain g , and its quantized value, when backward gain adaption is used. It can be seen that most of the codebook gain values have a magnitude less than or close to one, but it is still necessary to allocate two gain quantizer levels for the infrequently used high magnitude gain values.

By training a split 7/3 bit shape/gain codebook, as described in Section 12.7, for G728-like codecs both with and without gain adaption we found that the gain adaption increased the segmental SNR of the codec by 2.7 dB, and the weighted segmental SNR by 1.5 dB. These are very significant improvements, especially when it is considered

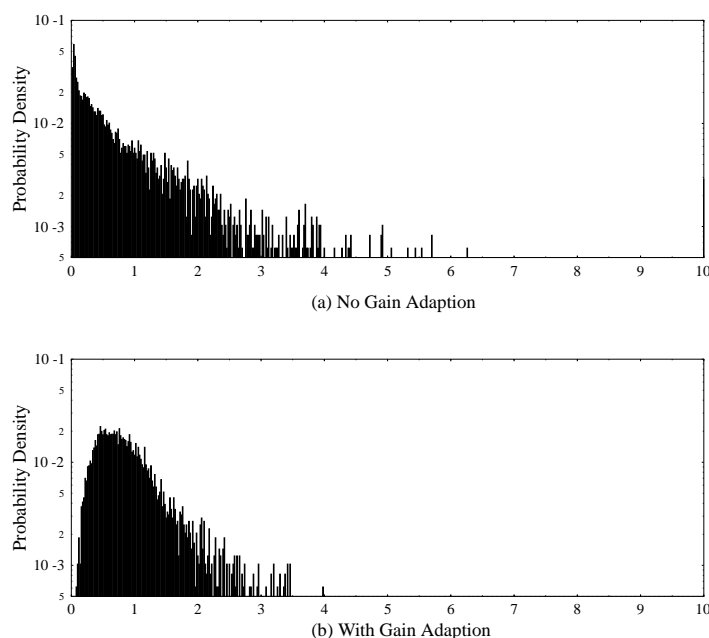


Figure 12.6: PDFs of the Normalised Codebook Gains With and Without Backward Gain Adaption

that the gain adaption increases the encoder complexity by only about 3%.

12.4.4 G728 Codebook Search

The standard recognised technique of finding the optimum excitation in CELP codecs is to generate the so-called **target vector** for each input speech vector to be encoded and match the filtered candidate excitation sequences to the target, as it will be explained in our forthcoming discourse. During synthesizing the speech signal for each codebook vector the excitation vectors are filtered through the concatenated LPC synthesis filter and the error weighting filter which are described by the impulse response $h(n)$ as seen in Figure 12.1. Since this filter complex is an infinite impulse response (IIR) system, upon exciting it with a new codebook entry its output signal will be the super-position of the response due to the current entry plus the response due to all previous entries. We note that the latter contribution is not influenced by the current input vector and hence this **filter memory contribution** plays no role in identifying the best codebook vector for the current 5-sample frame. Therefore the filter memory contribution due to previous inputs has to be buffered, before a new excitation is input and subtracted from the current input speech frame in order to generate the target vector $x(n)$, to which all filtered codebook entries are compared in order to find the best innovation sequence resulting in the best synthetic speech

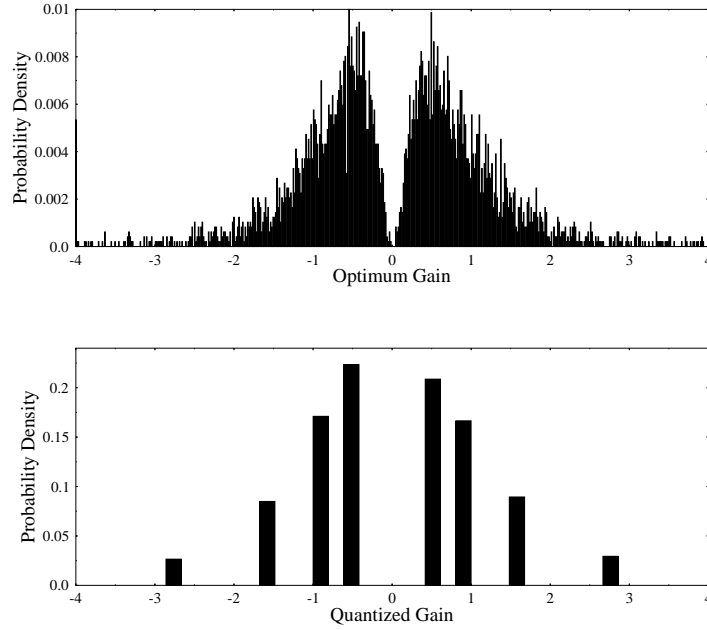


Figure 12.7: PDFs of the Optimum and Quantized Codebook Gain Values

segment. A preferred alternative to subtracting the filter memory from the input speech in generating the target vector is to set the filter memory to zero before a new codebook vector is fed into it. Since the backward adaptive gain $\sigma(n)$ is known at frame n , before the codebook search commences, the normalised target vector $x(n) = x(n)/\sigma(n)$ can be used during the optimisation process.

Let us follow the notation used in the G728 Recommendation and denote the codebook vectors by $y_j, j = 1...128$ and the associated gain factor by $g_i, i = 1...8$. Then the filtered and gain-scaled codebook vectors are given by the convolution:

$$\hat{x}_{ij} = \sigma(n) \cdot g_i[h(n) * y_j], \quad (12.13)$$

where again, $\sigma(n)$ represents the codebook gain determined by the backward-adaptive gain recovery scheme of Figure 12.5. By the help of the lower triangle convolution matrix of:

$$\mathbf{H} = \begin{bmatrix} h_0 & 0 & 0 & 0 & 0 \\ h_1 & h_0 & 0 & 0 & 0 \\ h_2 & h_1 & h_0 & 0 & 0 \\ h_3 & h_2 & h_1 & h_0 & 0 \\ h_4 & h_3 & h_2 & h_1 & h_0 \end{bmatrix} \quad (12.14)$$

Equation 12.13 can be expressed in a more terse form as follows:

$$\hat{x}_{ij} = \mathbf{H}\sigma(n)g_i y_j. \quad (12.15)$$

The best innovation sequence is deemed to be the one, which minimises the following *mse* distortion expression:

$$D = \|x(n) - \hat{x}_{ij}\|^2 = \sigma^2(n)\|\hat{x}(n) - g_i \cdot \mathbf{H}y_j\|^2 \quad (12.16)$$

where again $\hat{x}(n) = x(n)/\sigma(n)$ is the normalised target vector. Upon expanding the above term we arrive at:

$$D = \sigma^2(n) [\|\hat{x}(n)\|^2 - 2g_i \hat{x}^T \mathbf{H}y_j + g_i^2 \|\mathbf{H}y_j\|^2] \quad (12.17)$$

Since the normalised target vector energy $\|\hat{x}(n)\|^2$ and the codebook gain $\sigma(n)$ are constant for the duration of scanning the codebook, minimising D in Equation 8.36 is equivalent to:

$$\hat{D} = -2g_i \cdot p^T(n) \cdot y_j + g_i^2 E_j, \quad (12.18)$$

where the shorthand of $p(n) = \mathbf{H}^T \cdot \hat{x}(n)$ and $E_j = \|\mathbf{H}y_j\|^2$ was employed. Notice that E_j represents the energy of the filtered codebook entry y_j , and since the filter coefficients are only updated every 20 samples, $E_j j = 1 \dots 128$ is computed once per LPC update frame.

The optimum codebook entry can now be found by identifying the best $g_i, i = 1 \dots 8$. A computationally more efficient technique is to compute the optimum gain factor for each entry and then quantise it to the closest prestored value. Further specific details of the codebook search procedure are given in [217, 231], while the codebook training algorithm was detailed in [345].

In the original CELP codec proposed by Schroeder and Atal a stochastic codebook populated by zero-mean unit-variance Gaussian vector was used. The G728 codec uses a 128-entry trained codebook.

In a conceptually simplistic, but suboptimum approach the codebook could be trained by simply generating the prediction residual using a stochastic codebook and then employ the **pairwise nearest neighbour** or the **pruning method** [248] to cluster the excitation vectors in order to arrive at a trained codebook. It is plausible however that upon using this trained codebook the prediction residual vectors generated during the codec's future operation will be now different, necessitating the re-training of the codebook recursively a number of times. This is particularly true in case of backward adoptive gain recovery, because the gain factor will be dependent on the codebook entries, which in turn again will depend on the gain values. According to Chen [345] the codec performance is dramatically reduced, if no closed-loop training is invoked. The robustness against channel errors was substantially improved following the proposals by De Marca and Jayant [346] as well as Zeger and Gersho [347] using pseudo-Gray coding of the codebook indeces, which ensured that in case of a single channel error the corresponding codebook entry was similar to the original one.

12.4.5 G728 Excitation Vector Quantization

At 16 kbits/s there are 10 bits which can be used to represent every 5 sample vector, and as the LPC analysis is backward adaptive these bits are used entirely to code the excitation signal $u(n)$ which is fed to the synthesis filter. The 5 sample excitation sequences are vector quantized using a 10 bit split shape-gain codebook. Seven bits are used to represent the vector shapes, and the remaining 3 bits are used to quantize the vector gains. This splitting of the 10 bit vector quantizer is done to reduce the complexity of the closed-loop codebook search. To measure the degradations that were introduced by this splitting we trained codebooks for a 7/3 bit shape/gain split vector quantizer, and a pure 10 bit vector quantizer. We found that the 10 bit vector quantizer gave no significant improvement in either the segmental SNR or the segmental weighted SNR of the codec, and increased the complexity of the codebook search by about 550% and the overall codec complexity by about 300%. Hence this splitting of the vector quantizer is a very efficient way to significantly reduce the complexity of the encoder.

The closed-loop codebook search is carried out as follows. For each vector the search procedure finds values of the gain quantizer index i and the shape codebook index k which minimise the squared weighed error E_w for that vector. E_w is given by

$$E_w = \sum_{n=0}^{vs-1} (s_w(n) - \hat{s}_o(n) - \hat{\sigma} g_i h(n) * c_k(n))^2 \quad (12.19)$$

where $s_w(n)$ is the weighted input speech, $\hat{s}_o(n)$ is the zero-input response of the synthesis and weighting filters, $\hat{\sigma}$ is the predicted vector gain, $h(n)$ is the impulse response of the concatenated synthesis and weighting filters and g_i and $c_k(n)$ are the entries from the gain and shape codebooks. This equation can be expanded to give:

$$E_w(n) = \hat{\sigma}^2 \sum_{n=0}^{vs-1} (x(n) - g_i [h(n) * c_k(n)])^2 \quad (12.20)$$

$$\begin{aligned} &= \hat{\sigma}^2 \sum_{n=0}^{vs-1} x^2(n) + \hat{\sigma}^2 g_i^2 \sum_{n=0}^{vs-1} [h(n) * c_k(n)]^2 \\ &\quad - 2\hat{\sigma}^2 g_i \sum_{n=0}^{vs-1} x(n) [h(n) * c_k(n)] \end{aligned} \quad (12.21)$$

$$= \hat{\sigma}^2 \sum_{n=0}^{vs-1} x^2(n) + \hat{\sigma}^2 (g_i^2 \xi_k - 2g_i C_k)$$

where $x(n) = (s_w(n) - \hat{s}_o(n))/\hat{\sigma}$ is the codebook search target,

$$C_k = \sum_{n=0}^{vs-1} x(n) [h(n) * c_k(n)] \quad (12.22)$$

is the correlation between this target and the filtered codeword $h(n) * c_k(n)$, and

$$\xi_k = \sum_{n=0}^{vs-1} [h(n) * c_k(n)]^2 \quad (12.23)$$

is the energy of the filtered codeword $h(n) * c_k(n)$. Note that this is almost identical to the form of the term in Equation 10.7 which must be minimised in the fixed codebook search in our ACELP codecs.

In the G728 codec the synthesis and weighting filters are changed only once every four vectors. Hence ξ_k must be calculated for the 128 codebook entries only once every four vectors. The correlation term C_k can be rewritten as

$$\begin{aligned} C_k &= \sum_{n=0}^{vs-1} x(n)[h(n) * c_k(n)] \\ &= \sum_{n=0}^{vs-1} c_k(n)\psi(n) \end{aligned} \quad (12.24)$$

where

$$\psi(n) = \sum_{i=n}^{vs-1} x(i)h(i-n) \quad (12.25)$$

is the reverse convolution between $h(n)$ and $x(n)$. This means that we need to carry out only one convolution operation for each vector to find $\psi(n)$ and then we can find C_k for each codebook entry k with a relatively simple series of multiply-add operations.

The codebook search finds the codebook entries $i=1-8$ and $k=1-128$ which minimise E_w for the vector. This is equivalent to minimising

$$D_{ik} = g_i^2 \xi_k - 2g_i C_k. \quad (12.26)$$

For each codebook entry k , C_k is calculated and then the best quantized gain value g_i is found. The values g_i^2 and $2g_i$ are pre-computed and stored for the 8 quantized gains, and these values along with ξ_k and C_k are used to find D_{ik} . The codebook index k which minimises this, together with the corresponding gain quantizer level i , are sent to the decoder. These indices are also used in the encoder to produce the excitation and reconstructed speech signals which are used to update the gain predictor and the synthesis filter.

The decoder's schematic was portrayed in Figure 12.2, which carries out the inverse operations of the encoder seen in Figure 12.2. Without delving into specific algorithmic details of the decoder's functions in the next Section we briefly describe the operation of the postfilter at its output stage.

Post filtering was originally proposed by Jayant and Ramamoorthy, [228, 229] in the context of ADPCM coding using the two-pole six-zero synthesis filter of the G721/codec of Figure 6.11 to improve the perceptual speech quality.

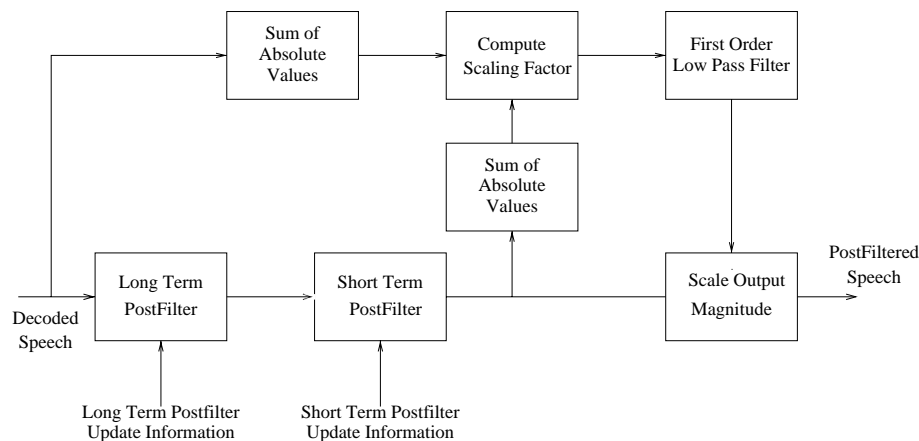


Figure 12.8: G728 Postfilter Schematic

12.4.6 G728 Adaptive Postfiltering

Since post-filtering was shown to improve the perceptual speech quality in the G721 ADPCM codec, Chen *et al.* have also adopted this technique in order to improve the performance of CELP codecs [230]. The basic philosophy of post-filtering is to augment spectral prominances, while slightly reducing their bandwidth and attenuating spectral valleys between them. This procedure naturally alters the waveform shape to a certain extent, which constitutes an impairment, but its perceptual advantage in terms of reducing the effect of quantisation noise outweighs the former disadvantage.

Early versions of the G728 codec did not employ adaptive post-filtering in order to prevent the accumulation of speech distortion during tandeming several codecs. However, without post-filtering the coding noise due to concatenating three asynchronously operated codecs became about 4.7 dB higher than in case of one codec. Chen *et al.* found that this was due to optimising the extent of post-filtering for maximum noise masking at a concomitant minimum speech distortion, while using a single coding stage. Hence the amount of post-filtering became excessive in case of tandeming. This then led to a design, which was optimised for three concatenated coding operation and the corresponding speech quality improved by a Mean Opinion Score (MOS) point of 0.81 to 3.93.

12.4.6.1 Adaptive Long-term Postfiltering

The schematic of the G728 adaptive post-filter is shown in Figure 12.8. The **long-term postfilter** is a comb filter, which enhances the spectral needles in the vicinity of the upper harmonics of the pitch frequency. Albeit the G728 codec dispenses with using LTP or pitch predictor for reasons of error resilience, the pitch information is recovered in the codec using a pitch detector to be described at a later stage. Assuming that the true pitch periodicity p is known, the long-term post-filter can be described by

the help of the transfer function:

$$H_l = g_l(1 + bz^{-p}), \quad (12.27)$$

where the coefficients g_l , b and p are updated during the third 5-sample speech segment of each 4-segment, or 2.5 ms duration LPC update frame, as suggested by Figure 12.8.

The **postfilter adapter** schematic is displayed in Figure 12.9. A 10-th order LPC inverse filter and the **pitch detector** act in unison in order to extract the pitch periodicity p . Chen *et al.* also proposed a possible implementation for the pitch detector. The 10-th order LPC inverse filter of

$$\tilde{A}(z) = 1 - \sum_{i=1}^{10} \tilde{a}_i z^{-i} \quad (12.28)$$

employs the filter coefficients \tilde{a}_i , $i = 1 \dots 10$ computed from the synthetic speech in order to generate prediction residual $d(k)$. This signal is fed to the pitch detector of Figure 12.9, which buffers a 240-sample history of $r(k)$. It would now be possible to determine the pitch periodicity using the straightforward evaluation of Equation 7.7 for all possible delays in the search scope, which was stipulated in the G78 codec to be [20 ... 140], employing a summation limit of $N=100$. However, the associated complexity would be unacceptably high.

Therefore the Recommendation suggests to low-pass filter $d(k)$ using a third-order elliptic filter to a bandwidth of 1kHz and then decimate it by a factor of four, allowing a substantial complexity reduction. The second term of Equation 7.7 is maximised over the search scope of $\alpha = [20, 21 \dots 140]$, but in the decimated domain this corresponds to the range [5, 6, ... 35]. Now Equation 7.7 only has to be evaluated for 31 different delays and the $\log \alpha_1$ maximising the second term of Equation 7.7 is inferred as an initial estimate of the true pitch periodicity p . This estimate can then be refined to derive a better estimate α_3 by maximising the above mentioned second term of Equation 7.7 over the undecimated $r(k)$ signal within the log range of $[\alpha_1 \pm 3]$. In order to extract the true pitch periodicity, it has to be established, whether the refined estimate α_2 is not a multiple of the true pitch. This can be ascertained by evaluating the second term of Equation 7.7 also in the range $[\alpha_3 \pm 6]$, where α_3 is the pitch determined during the previous 20-sample LPC update frame. Due to this frequent pitch-picking update at the beginning of each talk-spurt the scheme will be able to establish the true pitch lag, since the true pitch lag is always longer than 20 samples or 2.5 ms and hence no multiple-length lag values will be detected. This will allow the codec to recursively check in the absence of channel error, whether the current pitch lag is within a range of ± 6 samples or 1.5ms of the previous one, namely α_3 . If this is not the case, the lag $(\alpha_3 - 6) < \alpha_4 < (\alpha_3 + 6)$ is also found, for which the second term of Equation 7.7 is maximum.

Now a decision must be taken, as to whether α_4 or α_2 constitutes the true pitch lag and this can be established by ranking them on the basis of their associated gain terms $G = \beta$ given by Equation 7.6, which is physically the normalised cross-correlation of the residual segments at delays 0 and α , respectively. The higher this correlation, the more likely that α represents the true pitch lag. In possession of the optimum

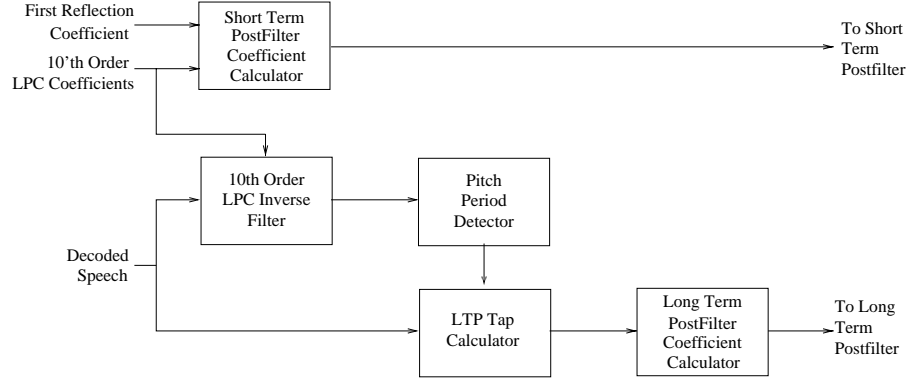


Figure 12.9: Postfilter adapter schematic

LTP lag α and gain β Chen *et al.* defined the LT postfilter coefficients b and g_e in Equation 12.27 as

$$b = \begin{cases} 0 & \text{if } \beta < 0.6 \\ 0.15\beta & \text{if } 0.6 \leq \beta \leq 1 \\ 0.15 & \text{if } \beta = 1 \end{cases} \quad (12.29)$$

$$g_e = \frac{1}{1 + b}, \quad (12.30)$$

where the factor 0.15 is an experimentally determined constant controlling the weighting of the LT postfilter. If the LTP gain of Equation 7.6 is close to unity, the signal $v(k)$ is almost perfectly periodic. If, however, $\beta < 0.6$, the signal is unvoiced, exhibiting almost no periodicity, hence the spectrum has no quasi-periodic fine-structure. Therefore according to $b = 0$ no long-term post-filtering is employed, since $H_l(z) = 1$ represents an all-pass filter. Lastly, in the range of $0.6 \leq \beta \leq 1$ we have $b = 0.5\beta$, ie, β controls the extent of long-term post-filtering, allowing a higher degree of weighting in case of highly correlated $r(k)$ and speech signals.

Having described the adaptive long-term post-filtering let us now turn our attention to details of the **short-term (ST) postfiltering**.

12.4.6.2 G728 Adaptive Short-term Postfiltering

The adaptive ST post-filter standardised in the G728 Recommendation is constituted by a 10-th order pole-zero filter concatenated with a first-order single-zero filter, as seen below:

$$H_s(z) = \frac{1 - \sum_{i=1}^{10} \bar{b}_i z^{-i}}{1 - \sum_{i=1}^{10} \bar{a}_i z^{-i}} [H_\mu z^{-1}], \quad (12.31)$$

where the filter coefficients are specified as follows:

$$\begin{aligned} \bar{b}_i &= \tilde{a}_i (0.65)^i & i = 1, 2, \dots, 10 \\ \bar{a}_i &= \tilde{a}_i (0.75)^i & i = 1, 2, \dots, 10 \end{aligned}$$

Synthesis Filter	5.1
Backward Gain Adaption	0.4
Weighting Filter	0.9
Codebook Search	6.0
Post Filtering	3.2
Total Encoder Complexity	12.4
Total Decoder Complexity	8.7

Table 12.4: Millions of Operations per Second Required by G728 Codec

$$\mu = 0.15.k_i \quad (12.32)$$

The coefficients $\tilde{a}_i, i = 1 \dots 10$ are obtained in the usual fashion as by-products of the 50-th order LPC analysis at iteration $i = 10$, while k_1 represents the first reflection coefficient in the Levinson-Durbin algorithm of Figure 6.3. Observe in Equation 12.32 that the coefficients \overline{a}_i and \overline{b}_i are derived from the progressively attenuated \tilde{a}_i coefficients. The pole-zero section of this filter emphasises the formant structure of the speech signal, while attenuating the frequency regions between formants. The single-zero section has a high-pass characteristic and was included in order to compensate for the low-pass nature or spectral delay of the pole-zero section.

Returning to Figure 12.8, observe that the output signal of the adaptive postfilter is scaled, in order for its input and output signals to have the same power. The sum of the postfilter's input and output samples is computed, the required scaling factor is calculated and low-pass filtered in order to smooth its fluctuation, before the output scaling takes place.

Here we conclude our discussions on the standard G728 16 kbps codec with a brief performance analysis, before we embark on contriving a range of programmable-rate 8-16 kbps codecs.

12.4.7 Complexity and Performance of the G728 Codec

In the previous sub-sections we have described the operation of the G728 codec. The associated implementational complexities of the various sections of the codec are shown in Table 12.4 in terms millions of arithmetic operations (mostly multiplies and adds) per second. The weighting filter and codebook search operations are carried out only by the encoder, which requires a total of about 12.4 million operations per second. The post filtering is carried out only by the decoder which requires about 8.7 million operations per second. The full duplex codec requires about 21 million operations per second.

We found that the codec gave an average segmental SNR of 20.1 dB, and an average weighted segmental SNR of 16.3 dB. The reconstructed speech was difficult to distinguish from the original, with no obvious degradations. In the next Section we discuss our attempts to modify the G728 codec in order to produce a variable bit rate 8-16 kbits/s codec which gives a graceful degradation in speech quality as the bit rate is reduced. Such a programmable-rate codec is useful in intelligent systems, where the transceiver may be reconfigured under network control, in order to invoke a higher or

lower speech quality mode of operation, or to assign more channel capacity to error correction coding in various traffic loading or wave propagation scenarios.

12.5 Reduced-Rate G728-Like Codec: Variable-length Excitation Vector

Having detailed the G728 codec in the previous Section we now describe our work in reducing the bit rate of this codec and producing an 8-16 kbits/s variable rate low-delay codec. The G728 codec uses 10 bits to represent each 5 sample vector. It is obvious that to reduce the bit rate of this codec we must either reduce the number of bits used for each vector, or increase the number of speech samples per vector. If we were to keep the vector size fixed at 5 samples then in an 8 kbits/s codec we would have only 5 bits to represent both the excitation shape and gain. Without special codebook training this leads to a codec with unacceptable performance. Therefore initially we concentrated on reducing the bit rate of the codec by increasing the vector size. In Section 12.8 we discuss the alternative approach of keeping the vector size constant and reducing the size of the codebooks used.

In this Section at all bit rates we use a split 7/3 bit shape/gain vector quantizer for the excitation signal $u(n)$. The codec rate is varied by changing the vector size vs used, from $vs = 5$ for the 16 bits/s codec to $vs = 10$ for the 8 kbits/s codec. For all the codecs we used the same 3 bit gain quantizer as in G728, and for the various shape codebooks we used randomly generated Gaussian codebooks with the same variance as the G728 shape codebook. Random codebooks with a Gaussian PDF were used for simplicity and because in the past such codebooks have been shown to give a relatively good performance [16]. We found that replacing the trained shape codebook in the G728 codec with a Gaussian codebook reduced the segmental SNR of the codec by 1.7 dB, and the segmental weighted SNR by 2 dB. However these losses in performance are recovered in Section 12.7 when we consider closed-loop training of our codebooks.

In the G728 codec the synthesis filter, weighting filter and the gain predictor are all updated every four vectors. With a vector size of 5 this means the filters are updated every 20 samples or 2.5ms. Generally the more frequently the filters are updated the better the codec will perform, and we found this to be true for our codec. However updating the filter coefficients more frequently significantly increases the complexity of the codec. Therefore we decided to keep the period between filter updates as close as possible to 20 samples as the bit rate of our codec is reduced by increasing the vector size. This means reducing the number of vectors between filter updates as the vector size is increased. For example at 8 kbits/s the vector size is 10 and we updated the filters every 2 vectors, which again corresponds to 2.5ms.

The segmental SNR of our codec against its bit rate as the vector size is increased from 5 to 10 is shown in Figure 12.10. Also shown in this figure is the segmental prediction gain of the synthesis filter at the various bit rates. It can be seen from this figure that the segmental SNR of our codec decreases smoothly as its bit rate is reduced, falling by about 0.8 dB for every 1 kbits/s drop in the bit rate.

As explained in the previous Section, an important part of the codec is the backward

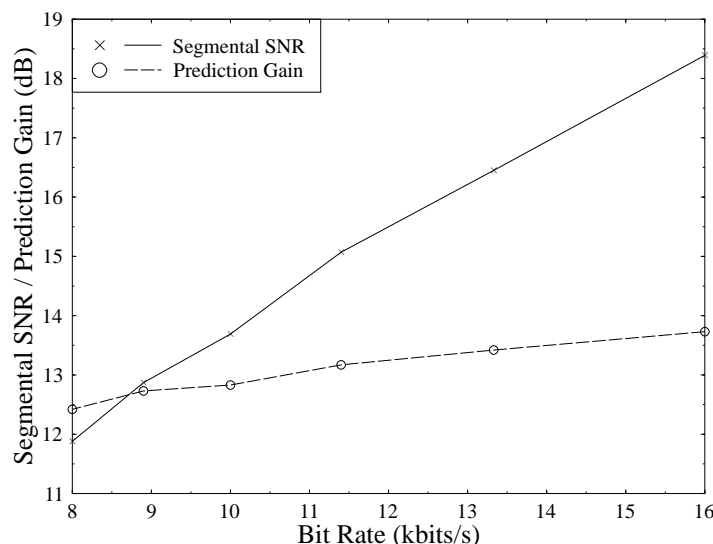


Figure 12.10: Performance of the Reduced-Rate G728-Like Codec I with Variable-length Excitation Vectors

adaptive synthesis filter. It can be seen from Figure 12.10 that the prediction gain of this filter falls by only 1.3 dB as the bit-rate of the codec is reduced from 16 to 8 kbits/s. This suggests that the backward adaptive synthesis filtering copes well with the reduction in bit rate from 16 to 8 kbits/s. We also carried out tests at 16 and 8 kbits/s, similar to those used for Table 12.3, to establish how the performance of the filter would be improved if we were able to eliminate the effects of using backward adaption ie the noise feedback and time mismatch. The results are shown in Tables 12.5 and 12.6 for the 16 kbits/s codec (using the Gaussian codebook rather than the trained G728 codebook used for Table 12.3) and the 8 kbits/s codec. As expected the effects of noise feedback are more significant at 8 than 16 kbits/s, but the overall effects on the codec's segmental SNR of using backward adaption are similar at both rates.

It has been suggested [339] that high order backward adaptive linear prediction is inappropriate at bit rates as low as 8 kbits/s. However we found that this was not the case for our codec and that increasing the filter order from 10 to 50 gave almost the same increase in the codec performance at 8 kbits/s as at 16 kbits/s. This is shown in Table 12.7.

Another important part of the G728 codec is the backward gain adaption. Figure 12.6 shows how at 16 kbits/s this backward adaption makes the optimum codebook gains cluster around one, and hence become easier to quantize. We found that the

	Δ Prediction Gain (dB)	Δ Segmental SNR (dB)
No Noise Feedback	+0.74	+0.42
No Time Mismatch	+0.85	+0.83
Use Forward Adaption	+1.59	+1.25

Table 12.5: Effects of Backward Adaption of the Synthesis Filter at 16 kbits/s

	Δ Prediction Gain (dB)	Δ Segmental SNR (dB)
No Noise Feedback	+2.04	+0.75
No Time Mismatch	+0.85	+0.53
Use Forward Adaption	+2.89	+1.28

Table 12.6: Effects of Backward Adaption of the Synthesis Filter at 8 kbits/s

same was true at 8 kbits/s. To quantify the performance of the gain prediction we defined the following signal to noise ratio

$$SNR_{\text{gain}} = \frac{\sum \sigma_o^2}{\sum (\sigma_o - \hat{\sigma})^2}. \quad (12.33)$$

Here σ_o is the optimum excitation gain given by

$$\sigma_o = \sqrt{\frac{1}{vs} \sum_{n=0}^{vs} (\hat{\sigma} g c_k(n))^2} \quad (12.34)$$

where g is the unquantized gain chosen by the codebook search and $\hat{\sigma}$ is the predicted gain value. We found that this gain prediction SNR was on average 5.3 dB for the 16 kbits/s codec, and 6.1 dB for the 8 kbits/s codec. Thus the gain prediction is even more effective at 8 kbits/s than at 16 kbits/s.

In the next Section we discuss the addition of long term prediction to our variable rate codec.

	Δ Prediction Gain (dB)	Δ Segmental SNR (dB)
8 kbits/s p=10	0.0	0.0
8 bits/s p=50	+0.88	+1.00
16 kbits/s p=10	0.0	0.0
16 kbits/s p=50	+1.03	+1.04

Table 12.7: Relative Performance of the Synthesis Filter as p is Increased at 8 and 16 kbits/s

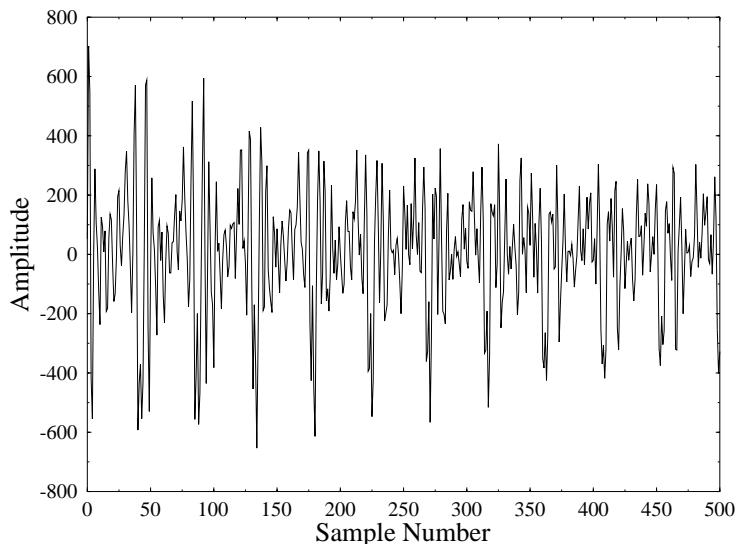


Figure 12.11: Short-term Synthesis-filter Prediction Residual in G728

12.6 The Effects of Long Term Prediction

In this Section we describe the improvements in our variable rate codec that can be obtained by adding backward adaptive Long Term Prediction (LTP). This work was motivated by the fact that we found significant long term correlations remained in the synthesis filter's prediction residual, even when the pitch period was lower than the order of this filter. This can be seen from Figure 12.11, which shows the prediction residual for a segment of voiced female speech with a pitch period of about 45 samples. It can be seen that the residual has clear long term redundancies, which could be exploited by a long term prediction filter.

In a forward adaptive system the short term synthesis filter coefficients are determined by minimising the energy of the residual signal found by filtering the original speech through the inverse synthesis filter. Similarly for open-loop LTP we minimise the energy of the long term residual signal which is found by filtering the short term residual through the inverse long term predictor. If $r(n)$ is the short term residual signal, then for a one tap long term predictor we want to determine the delay L and gain β which minimise the long term residual energy E_{LT} given by

$$E_{LT} = \sum_n (r(n) - \beta r(n - L))^2. \quad (12.35)$$

The best delay L is found by calculating

$$X = \frac{(\sum_n r(n)r(n-L))^2}{\sum_n r^2(n-L)} \quad (12.36)$$

for all possible delays, and choosing the value of L which maximises X . The best long term gain β is then given by

$$\beta = \frac{\sum_n r(n)r(n-L)}{\sum_n r^2(n-L)}. \quad (12.37)$$

In a backward adaptive system the original speech signal $s(n)$ is not available, and instead we use the past reconstructed speech signal $\hat{s}(n)$ to find the short term synthesis filter coefficients. These coefficients can then be used to filter $\hat{s}(n)$ through the inverse filter to find the “reconstructed residual” signal $\hat{r}(n)$. This residual signal can then be used in Equations 12.36 and 12.37 to find the LTP delay and gain. Alternatively we can use the past excitation signal $u(n)$ in Equations 12.36 and 12.37. This approach is slightly simpler than using the reconstructed residual signal because the inverse filtering of $\hat{s}(n)$ to find $\hat{r}(n)$ is not necessary, and we found in our codec that the two approaches gave almost identical results.

Initially we used a one tap LTP in our codec. The best delay L was found by maximising

$$X = \frac{\left(\sum_{n=-100}^{-1} u(n)u(n-L)\right)^2}{\sum_{n=-100}^{-1} u^2(n-L)} \quad (12.38)$$

over the range of delays 20 to 140 every frame. The LTP gain β was updated every vector by solving

$$\beta = \frac{\sum_{n=-100}^{-1} u(n)u(n-L)}{\sum_{n=-100}^{-1} u^2(n-L)}. \quad (12.39)$$

We found that this backward adaptive LTP improved the average segmental SNR of our codec by 0.6 dB at 16 kbits/s, and 0.1 dB at 8 kbits/s. However the calculation of X as given in Equation 12.38 for 120 different delays every frame dramatically increases the complexity of the codec. The denominator $\sum u^2(n-L)$ for delay L need not be calculated independently, but instead can be simply updated from the equivalent expression for delay $L-1$. Even so if the frame size is 20 samples then to calculate X for all delays increases both the encoder and the decoder complexity by almost 10 million arithmetic operations per second, which is clearly unacceptable.

Fortunately the G728 post-filter requires an estimate of the pitch period of the current frame. This is found by filtering the reconstructed speech signal through a tenth order short term prediction filter to find a reconstructed residual like signal. This signal is then low pass filtered with a cut-off frequency of 1 kHz and 4:1 decimated, which dramatically reduces the complexity of the pitch determination. The maximum value of the auto-correlation function of the decimated residual signal is then found to give an estimate τ_d of the pitch period. A more accurate estimate τ_p is then found by maximising the autocorrelation function of the undecimated residual between $\tau_d - 3$

	Segmental SNR (dB)	Segmental Weighted SNR (dB)
No LTP	18.43	14.30
1 Tap LTP	19.08	14.85
3 Tap LTP	19.39	15.21
5 Tap LTP	19.31	15.12

Table 12.8: Performance of LTP at 16 kbits/s

	Segmental SNR (dB)	Segmental Weighted SNR (dB)
No LTP	11.86	8.34
1 Tap LTP	12.33	8.64
3 Tap LTP	12.74	9.02
5 Tap LTP	12.49	8.81

Table 12.9: Performance of LTP at 8 kbits/s

and $\tau_d + 3$. This lag could be a multiple of the true pitch period, and to guard against this possibility the autocorrelation function is also maximised between $\tau_o - 6$ and $\tau_o + 6$, where τ_o is the pitch period from the previous frame. Finally the pitch estimator chooses between τ_p and the best lag around τ_o by comparing the optimal tap weights β for these two delays.

This pitch estimation procedure requires only about 2.6 million arithmetic operations per second, and is carried out at the decoder as part of the post-filtering operations anyway. So using this method to find a LTP delay has no effect on the decoder complexity, and increases the encoder complexity by only 2.6 million arithmetic operations per second. We also found that not only was this method of calculating the LTP delay much simpler than finding the maximum value of X from Equation 12.38 for all delays between 20 and 140, it also gave better results. This was due to the removal of pitch doubling and tripling by the checking of pitch values around that used in the previous frame. The average segmental SNR and segmental weighted SNR for our codec at 16 kbits/s both with and without one tap LTP using the pitch estimate from the post-filter is shown in Table 12.8. Similar figures for the codec at 8 kbits/s are given in Table 12.9. We found that when LTP was used, there was very little gain in having a filter order any higher than 20. Therefore the figures in Tables 12.8 and 12.9 have a short term filter order of 20 when LTP is used.

Tables 12.8 and 12.9 also give the performance of our codec at 16 and 8 kbits/s when we use multi-tap LTP. As the LTP is backward adaptive we can use as many taps in the filter as we like, with the only penalty being a slight increase in complexity. Once the delay is known, for a $(2p + 1)$ 'th order predictor the filter coefficients $b_{-p}, b_{-p+1}, \dots, b_0, \dots, b_p$ are given by solving the following set of simultaneous equa-

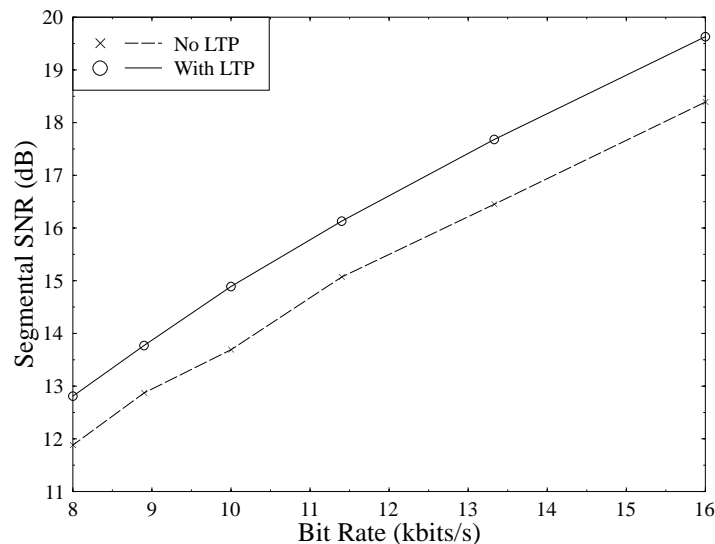


Figure 12.12: Performance of a 8-16 kbits/s Low Delay Codec With LTP

tions

$$\sum_{j=-p}^{j=p} b_j \sum_{n=-100}^{n=-1} u(n-L-j)u(n-L-i) = \sum_{n=-100}^{-1} u(n)u(n-L-i) \quad (12.40)$$

for $i = -p, -p+1, \dots, p$. The LTP synthesis filter $H_{LTP}(z)$ is then given by

$$H_{LTP}(z) = \frac{1}{1 - b_{-p}z^{-L+p} - \dots - b_0z^{-L} - \dots - b_pz^{-L-p}}. \quad (12.41)$$

It can be seen from Tables 12.8 and 12.9 that at both 16 and 8 kbits/s the best performance is given by a 3 tap filter which improves the segmental SNR at both bit rates by almost 1 dB. Also because when LTP is used the short term synthesis filter order was reduced to 20, the complexity of the codecs is not significantly increased by the use of a long term prediction filter.

We found that it was possible to slightly increase the performance of the codec with LTP by modifying the signal $u(n)$ used to find the filter coefficients in Equation 12.40. This modification involves simply repeating the previous vector's excitation signal once. Hence instead of using the signal $u(-1), u(-2), \dots, u(-100)$ to find the LTP coefficients, we use $u(-1), u(-2), \dots, u(-vs), u(-1), u(-2), \dots, u(-100+vs)$. This single repetition of the previous vector's excitation in the calculation of the

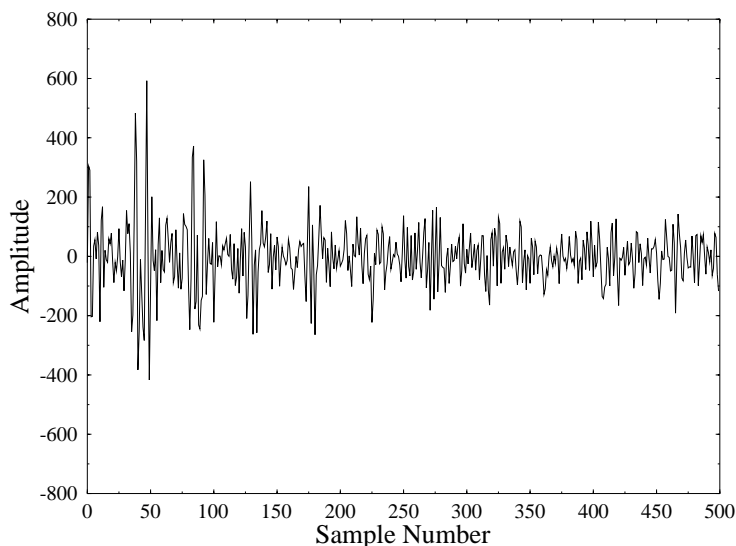


Figure 12.13: Long Term Filter Prediction Residual at 16 kbits/s

LTP coefficients increased both the segmental and the weighted SNR of our codec at 16 kbits/s by about 0.25 dB. It also improved the codec performance at 8 kbits/s, although only by about 0.1 dB. The improvements that this repetition brings in the codec's performance seem to be due to the backward adaptive nature of the LTP - no such improvement is seen when a similar repetition is used in a forward adaptive system.

Shown in Figure 12.12 is the variation in the codec's segmental SNR as the bit rate is reduced from 16 to 8 kbits/s. The codec uses 3 tap LTP with the repetition scheme described above and a short term synthesis filter of order 20. Also shown in this figure is the equivalent variation in segmental SNR for the codec without LTP, repeated here from Figure 12.10. It can be seen that the addition of long term prediction to the codec gives a uniform improvement in its segmental SNR of about 1 dB from 8 to 16 kbits/s. The effectiveness of the LTP can also be seen from Figure 12.13 which shows the long term prediction residual, in the 16 kbits/s codec, for the same segment of speech as was used for the short term prediction residual in Figure 12.11. It is clear that the long term correlations have been significantly reduced. It should be noted however that the addition of backward adapted long term prediction to the codec will degrade its performance over noisy channels. This aspect of our codec's performance is the subject of ongoing work [110].

Finally we tested the degradations in the performance of the long term prediction due to backward adaption being used. To measure the effect of quantization noise

	Δ Segmental Weighted SNR (dB)	Δ Segmental SNR (dB)
No Noise Feedback	-0.03	+0.01
No Time Mismatch	+0.87	+0.85
Use Forward Adaption	+0.84	+0.86

Table 12.10: Effects of Backward Adaption of the LTP at 16 kbits/s

	Δ Segmental Weighted SNR (dB)	Δ Segmental SNR (dB)
No Noise Feedback	-0.18	+0.02
No Time Mismatch	+1.17	+1.17
Use Forward Adaption	+0.99	+1.19

Table 12.11: Effects of Backward Adaption of the LTP at 8 kbits/s

feedback we used past values of the original speech signal rather than the reconstructed speech signal to find the LTP delay and coefficients. To measure the overall effect of backward adaption as opposed to open-loop forward adaption we used both past and present speech samples to find the LTP delay and coefficients. The improvements obtained in terms of the segmental SNR and the segmental weighted SNR are shown in Table 12.10 for the codec at 16 kbits/s, and Table 12.11 for the codec at 8 kbits/s. It can be seen that the use of backward adaption degrades the codecs performance by just under 1 dB at 16 kbits/s, and just over 1 dB at 8 kbits/s. At both bit rates noise feedback has very little effect, with most of the degradation coming from the time mismatch inherent in backward adaption.

12.7 Closed-Loop Codebook Training

In this Section we describe the training of the shape and gain codebooks used in our codec at its various bit rates. In Sections 12.5 and 12.6 Gaussian shape codebooks were used, together with the G728 gain codebook. These codebooks were used for simplicity, and in order to provide a fair comparison between the different coding techniques used.

Due to the backward adaptive nature of the gain and synthesis filter and LTP adaption used in our codec it is not sufficient to generate a training sequence for the codebooks and use the Lloyd algorithm [348] to design the codebooks. This is because the codebook entries required from the shape and gain codebooks depend very much upon the effectiveness of the gain adaption and the LTP and synthesis filters used. However, because these are backward adapted, they depend on the codebook entries that have been selected in the past. Therefore the effective training sequence needed changes as the the codebooks are trained. Thus it is reported in [342] for example that in a gain-adaptive vector quantization scheme unless the codebook is properly designed, taking into account the gain adaption, the performance is worse than simple non-adaptive vector quantization.

We used a closed-loop codebook design algorithm similar to that described in [345]. A long speech file consisting of four sentences spoken by two males and two females is used for the training. Both the sentences spoken and the speakers are different from those used for the performance figures quoted in this chapter. The training process commences with an initial shape and gain codebook and codes the training speech as usual. The total weighted error E_k from all the vectors that used the codebook entry $c_k(n)$ is then given by

$$E_k = \sum_{m \in N_k} \left(\hat{\sigma}_m^2 \sum_{n=0}^{vs-1} (x_m(n) - g_m[h_m(n) * c_k(n)])^2 \right) \quad (12.42)$$

where N_k is the set of vectors that use $c_k(n)$, $\hat{\sigma}_m$ is the backward adapted gain for vector m , g_m is the gain codebook entry selected for vector m and $h_m(n)$ is the impulse response of the concatenated weighting filter and the backward adapted synthesis filter used in vector m . Finally $x_m(n)$ is the codebook target for vector m , which with $(2p+1)$ 'th order LTP is given by

$$x_m(n) = \frac{s_{wm}(n) - \hat{s}_{om}(n) - \sum_{j=-p}^{j=p} b_{jm} u_m(n - L_m - j)}{\hat{\sigma}_m}. \quad (12.43)$$

Here $s_{wm}(n)$ is the weighted input speech in vector m , $\hat{s}_{om}(n)$ is the zero input response of the weighting and synthesis filters, $u_m(n)$ is the previous excitation and L_m and b_{jm} are the backward adapted LTP delay and coefficients in vector m .

Equation 12.42 giving E_k can be expanded to yield:

$$\begin{aligned} E_k &= \sum_{m \in N_k} \left(\hat{\sigma}_m^2 \sum_{n=0}^{vs-1} (x_m(n) - g_m[h_m(n) * c_k(n)])^2 \right) \\ &= \sum_{m \in N_k} \left(\hat{\sigma}_m^2 \sum_{n=0}^{vs-1} x_m^2(n) + \hat{\sigma}_m^2 g_m^2 \sum_{n=0}^{vs-1} [h_m(n) * c_k(n)]^2 \right. \\ &\quad \left. - 2\hat{\sigma}_m^2 g_m \sum_{n=0}^{vs-1} x_m(n)[h_m(n) * c_k(n)] \right) \\ &= \sum_{m \in N_k} \left(\hat{\sigma}_m^2 \sum_{n=0}^{vs-1} x_m^2(n) + \hat{\sigma}_m^2 g_m^2 \sum_{n=0}^{vs-1} [h_m(n) * c_k(n)]^2 \right. \\ &\quad \left. - 2\hat{\sigma}_m^2 g_m \sum_{n=0}^{vs-1} p_m(n)c_k(n) \right), \end{aligned} \quad (12.44)$$

where $p_m(j)$ is the reverse convolution between $h_m(n)$ and the target $x_m(n)$. This expression can be partially differentiated with respect to element $n = j$ of the codebook

entry $c_k(n)$ to give

$$\frac{\partial E_k}{\partial c_k(j)} = \sum_{m \in N_k} \left(2\hat{\sigma}_m^2 g_m^2 \sum_{n=0}^{vs-1} c_k(n) H_m(n, j) - 2\hat{\sigma}_m^2 g_m p_m(j) \right) \quad (12.45)$$

where $H_m(n, j)$ is the autocorrelation of the delayed impulse response $h_m(n)$ and is given by

$$H_m(i, j) = \sum_{n=0}^{vs-1} h_m(n-i) h_m(n-j). \quad (12.46)$$

Setting these partial derivatives to zero gives the optimum codebook entry $c_k^*(n)$ for the cluster of vectors N_k as the solution of the set of simultaneous equations

$$\sum_{m \in N_k} \left(\hat{\sigma}_m^2 g_m^2 \sum_{n=0}^{vs-1} c_k^*(n) H_m(n, j) \right) = \sum_{m \in N_k} (\hat{\sigma}_m^2 g_m p_m(j)) \quad (12.47)$$

for $j = 0, 1, \dots, vs-1$.

A similar expression for the total weighted error E_i from all the vectors that use the gain codebook entry g_i is

$$\begin{aligned} E_i &= \sum_{m \in N_i} \left(\hat{\sigma}_m^2 \sum_{n=0}^{vs-1} (x_m(n) - g_i [h_m(n) * c_m(n)])^2 \right) \\ &= \sum_{m \in N_i} \left(\hat{\sigma}_m^2 \sum_{n=0}^{vs-1} x_m^2(n) + g_i^2 \hat{\sigma}_m^2 \sum_{n=0}^{vs-1} [h_m(n) * c_m(n)]^2 \right. \\ &\quad \left. - 2g_i \hat{\sigma}_m^2 \sum_{n=0}^{vs-1} x_m(n) [h_m(n) * c_m(n)] \right) \end{aligned} \quad (12.48)$$

where N_i is the set of vectors that use the gain codebook entry g_i , and $c_m(n)$ is the shape codebook entry used by the m 'th vector. Differentiating this expression with respect to g_i gives:

$$\begin{aligned} \frac{\partial E_i}{\partial g_i} &= \sum_{m \in N_i} \left(2g_i \hat{\sigma}_m^2 \sum_{n=0}^{vs-1} [h_m(n) * c_m(n)]^2 \right. \\ &\quad \left. - 2\hat{\sigma}_m^2 \sum_{n=0}^{vs-1} x_m(n) [h_m(n) * c_m(n)] \right) \end{aligned} \quad (12.49)$$

and setting this partial derivative to zero gives the optimum gain codebook entry g_i^* for the cluster of vectors N_i as:

$$g_i^* = \frac{\sum_{m \in N_i} \left(\hat{\sigma}_m^2 \sum_{n=0}^{vs-1} x_m(n) [h_m(n) * c_m(n)] \right)}{\sum_{m \in N_i} \left(\hat{\sigma}_m^2 \sum_{n=0}^{vs-1} [c_m(n) * h_m(n)]^2 \right)}, \quad (12.50)$$

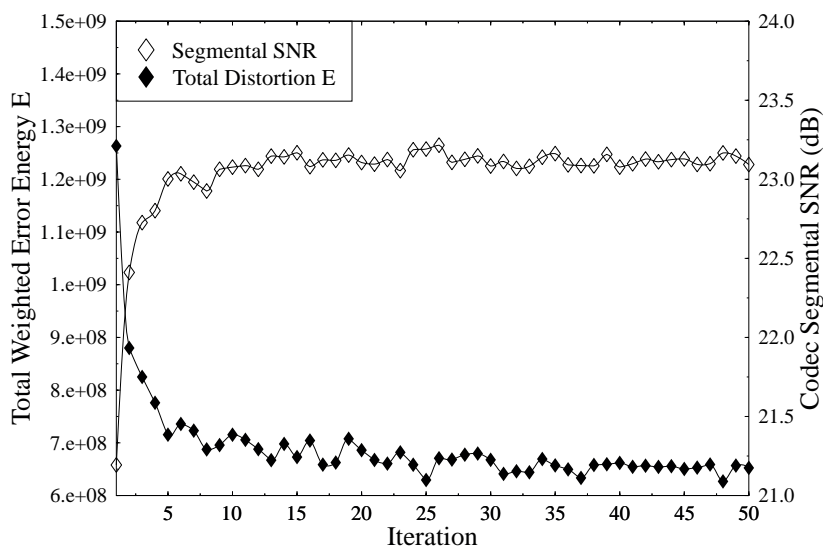


Figure 12.14: Codec's Performance as the Codebooks are Trained

The summations in Equations 12.48 and 12.50 over all the vectors that use $c_k(n)$ or g_i are carried out for all 128 shape codebook entries and all 8 gain codebook entries as the coding of the training speech takes place. At the end of the coding the shape and gain codebooks are updated using Equations 12.48 and 12.50, and then the codec starts coding the training speech again with the new codebooks. This closed loop codebook training procedure is summarised below

1. Start with an initial gain and shape codebook.
2. Code the training sequence using the given codebooks. Accumulate the summations in Equations 12.48 and 12.50.
3. Calculate the total weighted error of the coded speech. If this distortion is less than the minimum distortion so far keep a record of the codebooks used as the best codebooks so far.
4. Calculate new shape and gain codebooks using Equations 12.48 and 12.50.
5. Return to step 2.

Each entire coding of the training speech file counts as one iteration, and Figure 12.14 shows the variation in the total weighted error energy E , and the codec's segmental SNR, as the training progresses for the 16 kbits/s codebooks. From this figure

it can be seen that this closed-loop training sequence does not give a monotonic decrease in the total weighted error from one iteration to the next. This is because of the changing of the codebook target $x_m(n)$, as well as the other backward adapted parameters, from one iteration to the next. However it is clear from Figure 12.14 that the training does give a significant improvement in the codec's performance. Due to the non-monotonic decrease in the total weighted error energy it is necessary during the codebook training to keep a record of the lowest error energy achieved so far, and the corresponding codebooks. If a certain number of iterations passes without this minimum energy being improved then the codebook training can be terminated. It can be seen from Figure 12.14 that we get close to the minimum within about 20 iterations.

An important aspect in vector quantizer training can be the initial codebook used. In Figure 12.14 we used the G728 gain codebook and the Gaussian shape codebook as the initial codebooks. We also tried using other codebooks such as the G728 fixed codebook, and Gaussian codebooks with different variances, as the initial codebooks. However, although these gave very different starting values of the total weighted error E , and took different numbers of iterations to give their optimum codebooks, they all resulted in codebooks which gave very similar performances. Therefore we concluded that the G728 gain codebook, and the Gaussian shape codebook, are suitable for use as the initial codebooks.

We trained different shape and gain codebooks for use by our codec at all of its bit rates between 8 and 16 kbits/s. The average segmental SNR given by the codec using these codebooks is shown in Figure 12.15 for the 4 speech sentences which were not part of the training sequence. Also shown in this figure for comparison is the curve from Figure 12.12 for the corresponding codec with the untrained codebooks. It can be seen that the codebook training gives an improvement of about 1.5 to 2 dB across the codec's range of bit rates.

It can be seen from Figure 12.14 that a decrease in the total weighted error energy E does not necessarily correspond to an increase in the codec's segmental SNR. This is also true for the codec's segmental weighted SNR, and is because the distortion D calculated takes no account of the different signal energies in different vectors. We tried altering the codebook training algorithm described above to take account of this, hoping that it would result in codebooks which gave lower segmental SNRs. However the codebooks trained with this modified algorithm gave very similar performances to those trained by minimising E .

We also attempted training different codebooks at each bit rate for voiced and unvoiced speech. The voicing decision can be made backward adaptive based on the correlations in the previous reconstructed speech. A voiced/unvoiced decision like this is made in the G728 post-filter to determine whether to apply pitch post-filtering. We found however that although an accurate determination of the voicing of the speech could be made in a backward adaptive manner, no significant improvement in the codec's performance could be achieved by using separately trained voiced and unvoiced codebooks. This agrees with the results in [339] when fully backward adaptive LTP is used.

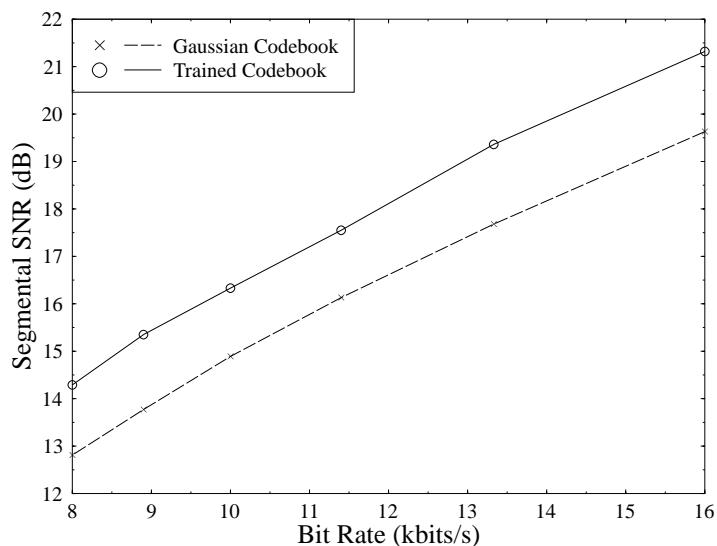


Figure 12.15: Performance of the 8-16 kbits/s Codec with Trained Codebooks

12.8 Reduced-Rate G728-Like Codec: Constant-length Excitation Vector

In the previous Sections we discussed a variable rate codec based on G728 which varied its bit rate by changing the number of samples in each vector. The excitation for each vector was coded with 10 bits. In this Section we describe the alternative approach of keeping the vector size constant and varying the number of bits used to code the excitation. The bit rate of the codec is varied between 8 and 16 kbits/s with a constant vector size of 5 samples by using between 5 and 10 bits to code the excitation signal for each vector. We used a structure for the codec identical to that described earlier, with backward gain adaption for the excitation and backward adapted short and long term synthesis filters. With 10,9 or 8 bits to code the excitation we used a split vector quantizer, similar to that used in G728, with a 7 bit shape codebook and a 3,2 or 1 bit gain codebook. For the lower bit rates we used a single 7,6 or 5 bit vector quantizer to code the excitation. Codebooks were trained for the various bit rates using the closed loop codebook training technique described in Section 12.7.

The segmental SNR of this variable rate codec is shown in Figure 12.16. Also shown in this graph is the segmental SNR of the codec with a variable vector size, copied here from Figure 12.15 for comparison. At 16 kbits/s the two codecs are of course identical, but at lower rates the constant vector size codec performs worse than the variable vector size codec. The difference between the two approaches increases as

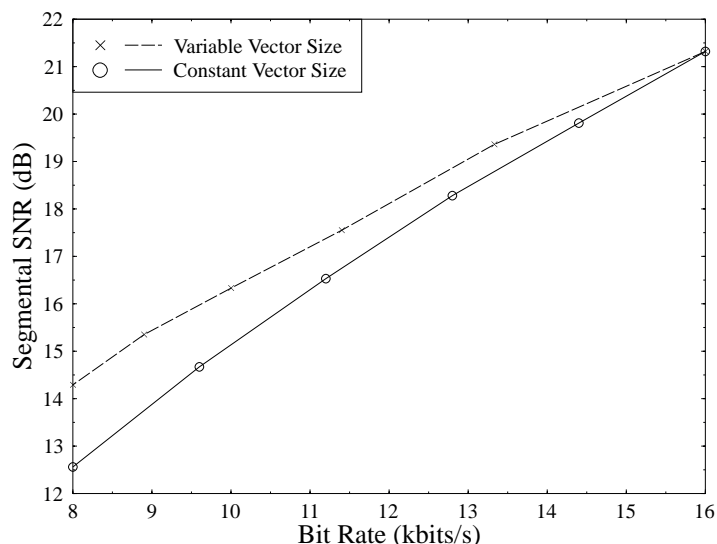


Figure 12.16: Performance of the Reduced-Rate G728-Like Codec II with Constant-length Excitation Vectors

the bit rate decreases, and at 8 kbits/s the segmental SNR of the constant vector size codec is about 1.75 dB lower than that of the variable vector size codec.

However, although the constant vector size codec gives lower reconstructed speech quality, it does have certain advantages. The most obvious is that it has a constant delay equal to that of G728, ie less than 2ms. Also the complexity of its encoder, especially at low bit rates, is lower than that of the variable vector size codec. This is because of the smaller codebooks used - at 8 kbits/s the codebook search procedure has only to examine 32 codebook entries. Therefore for some applications this codec may be more suitable than the higher speech quality variable vector size codec.

In this chapter so far we have described the G728 16 kbps low-delay codec and investigated a variable rate low delay codec, which is compatible with the 16 kbits/s G728 codec at its highest bit rate, and exhibits a graceful degradation in speech quality down to 8 kbits/s. The bit rate can be reduced while the buffering delay is kept constant at 5 samples (0.625 ms), or alternatively better speech quality is achieved if the buffering delay is increased gradually to 10 samples as the bit rate is reduced down to 8 kbits/s.

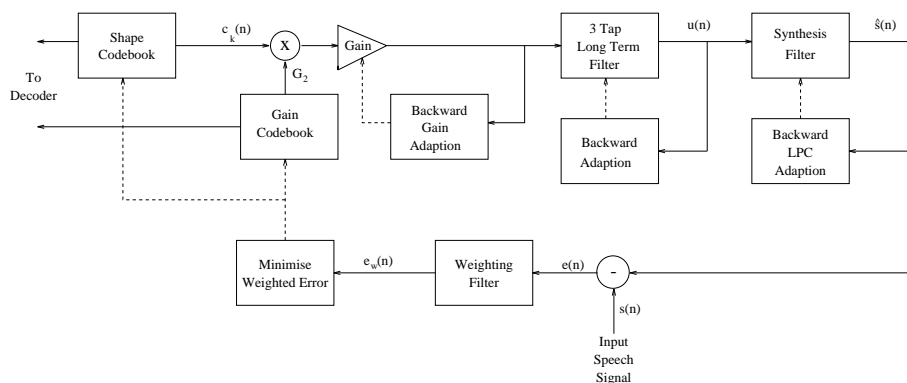


Figure 12.17: Scheme One Low Delay CELP Codec

12.9 Programmable-Rate 8-4 kbps Low Delay CELP Codecs

12.9.1 Motivation

Having discussed low delay 16-8 kbits/s programmable-rate coding in the previous Section, in this Section we consider methods of improving the performance of the proposed 8 kbits/s backward-adaptive predictive codec, while maintaining as low a delay and complexity as possible. Our proposed 8 kbits/s codec developed in Sections 12.5 and 12.8 uses a 3 bit gain codebook and a 7 bit shape codebook with backward adaption of both the long and the short term synthesis filters, and gives an average segmental SNR of 14.29 dB. In Section 12.9.2 we describe the effect of increasing the size of the gain and shape codebooks in this codec while keeping a vector length of 10 samples. This is followed by Sections 12.9.3 and 12.9.4 where we consider the improvements that can be achieved, again while maintaining a vector length of 10 samples, by using forward adaption of the short and long term synthesis filters. Then in Section 12.9.5 we show the performance of three codecs, based on those developed in the earlier Sections, operating at bit rates between 8 and 4 kbits/s. Finally, as an interesting benchmark, in Section 12.9.6 we describe a codec, with a vector size of 40 samples, based on the algebraic codebook structure we described in Section 10.4.3. The performance of this codec is compared to the previously introduced low delay codecs from Section 12.9.5 and the higher delay forward-adaptive predictive ACELP codec described in Section 10.4.3.

12.9.2 8-4kbps Codec Improvements Due to Increasing Codebook Sizes

In this Section we use the same structure for the codec as before, but increase the size of the shape and the gain codebooks. This codec structure is shown in Figure 12.17, and we refer to it as “Scheme One”. We used 3 tap backward adapted LTP and a vector length of 10 samples with a 7 bit shape codebook, and varied the size of

Gain Codebook Bits	Shape Codebook Bits	Segmental SNR (dB)
3	7	14.29
4	7	15.24
5	7	15.62
3	8	15.33
3	9	16.12
4	8	16.01

Table 12.12: Performance of the Scheme One Codec with Various Size Gain and Shape Codebooks

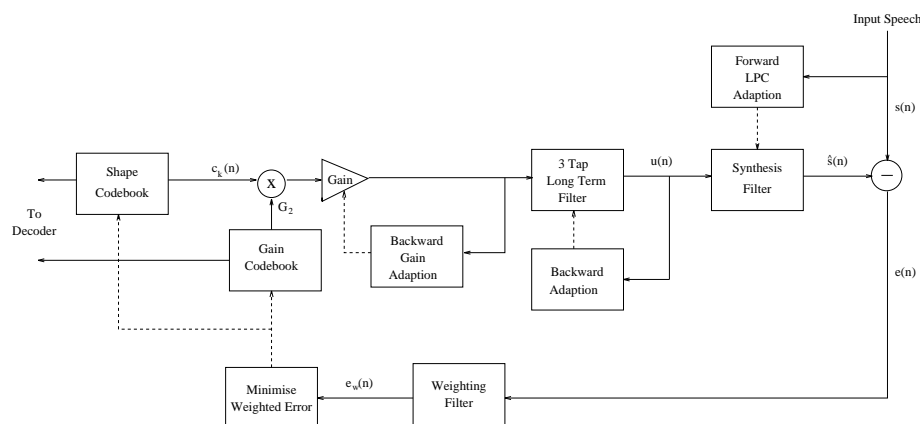


Figure 12.18: Scheme Two Low Delay CELP Codec

the gain codebook from 3 to 4 and 5 bits. Then in our next experiments we used a 3 bit gain codebook and trained 8 and 9 bit shape codebooks. Finally we attempted increasing the size of both the shape and the gain codebooks by one bit. In each case the new codebooks were closed-loop trained using the technique described in Section 12.7.

The segmental SNRs of this Scheme One codec with various size shape and gain codebooks is shown in Table 12.12. It can be seen that adding one bit to either the gain or the shape codebook increases the segmental SNR of the codec by about 1 dB. Adding two extra bits to the shape codebook, or one bit each to both codebooks, increases the segmental SNR by almost 2 dB.

12.9.3 8-4kbps Codecs - Forward Adaption of the Short Term Synthesis Filter

In this Section we consider the improvements that can be achieved in the vector size 10 codec by using forward adaption of the short term synthesis filter. In Table 12.6 we examined the effects of backward adaption of the synthesis filter at 8 kbits/s. However these figures gave the improvements that can be achieved by eliminating the

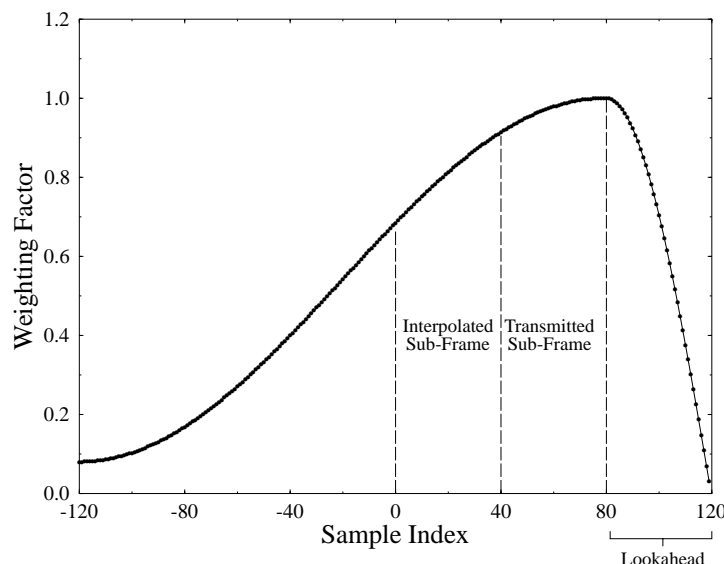


Figure 12.19: LPC Windowing Function Used in Candidate CCITT 8 kbits/s Codec

noise feedback and time mismatch that are inherent in backward adaption when using the same recursive windowing function and update rate as the G728 codec. In this Section we consider the improvements that could be achieved by significantly altering the structure used for the determination of the synthesis filter parameters.

The codec structure used is shown in Figure 12.18, and we refer to it as “Scheme Two”. Its only difference from our previously developed 8 kbits/s backward-adaptive codec is that we replaced the recursive windowing function shown in Figure 12.3 with an asymmetric analysis window which was used in a candidate codec for the CCITT 8 kbits/s standard [143, 322]. This window, which is shown in Figure 12.19, is made up of half a Hamming window and a quarter of a cosine function cycle. The windowing scheme uses a frame length of 10 ms (or 80 samples), with a 5 ms lookahead. The 10 ms frame consists of two sub-frames, and a Line Spectral Frequency (LSF) interpolation scheme similar to that described in Section 10.4.3 is used.

We implemented this method of deriving the LPC coefficients in our codec. The vector length was kept constant at 10 samples, but instead of the synthesis filter parameters being updated every 20 samples, as in the Scheme One codec, they were updated every 40 samples using either the interpolated or transmitted LSFs. In the candidate 8 kbits/s CCITT codec [143] a filter order of ten is used and the ten LSFs are quantized with 19 bits using differential split vector quantization. However for simplicity, and in order to see the best performance gain possible for our codec by using forward adaption of the short term synthesis filter, we used the ten unquantized LSFs

to derive the filter coefficients. A new 3 bit gain codebook and 7 bit shape codebook were derived for this codec using the codebook training technique described in Section 12.7. We found that this forward adaption increased the segmental SNR of the codec by only 0.8 dB, and even this rather small improvement would of course be reduced by the quantization of the LSFs. Using a 19 bit quantization scheme to transmit a new set of LSFs every 80 sample frame would mean using on average about 2.4 bits per 10 sample vector.

Traditionally codecs employing forward adaptive LPC are more resilient to channel errors than those using backward adaptive LPC. However a big disadvantage of using such a forward adaptive LPC scheme is that it would increase the delay of the codec by almost an order of magnitude. Instead of a vector length of 10 samples we would need to buffer a frame of 80 speech samples, plus a 40 sample look-ahead, to calculate the LPC information. This would increase the overall delay of the codec from under 4 ms to about 35 ms.

12.9.4 Forward Adaption of the Long Term Predictor

12.9.4.1 Initial Experiments

In this Section we consider the gains in our codec performance which can be achieved using forward adaption of the Long Term Predictor (LTP) gain. Although forward adaption of the LTP parameters would improve the codec's robustness to channel errors, we did not consider forward adaption of the LTP delay because to transmit this delay from the encoder to the decoder would require around 7 extra bits per vector. However we expected to be able to improve the performance of the codec, at the cost of significantly fewer extra bits, by using forward adaption of the LTP gain.

Previously we employed a three-tap LTP with backward adapted values for the delay and filter coefficients. Initially we replaced this LTP scheme with an adaptive codebook arrangement, where the delay was still backward adapted but the gain was calculated as in forward-adaptive CELP codecs, which was detailed in Section 10.5. This calculation assumes that the fixed codebook signal, which is not known until after the LTP parameters are calculated, is zero. The "optimum" adaptive codebook gain G_1 , which minimises the weighted error between the original and reconstructed speech is then given by:

$$G_1 = \frac{\sum_{n=0}^{vs-1} x(n)y_\alpha(n)}{\sum_{n=0}^{vs-1} y_\alpha^2(n)}. \quad (12.51)$$

Here $x(n) = s_w(n) - \hat{s}_o(n)$ is the target for the adaptive codebook search, $s_w(n)$ is the weighted speech signal, $\hat{s}_o(n)$ is the zero input response of the weighted synthesis filter, and

$$y_\alpha(n) = \sum_{i=0}^n u(i - \alpha)h(n - i) \quad (12.52)$$

is the convolution of the adaptive codebook signal $u(n - \alpha)$ with the impulse response $h(n)$ of the weighted synthesis filter, where α is the backward adapted LTP delay.

Again, we trained new 7/3-bit shape/gain fixed codebooks, and used the unquantized LTP gain G_1 as given by Equation 12.51. However we found that this arrange-

ment improved the segmental SNR of our codec by only 0.1 dB over the codec with 3 tap backward adapted LTP. Therefore we decided to invoke some of the joint adaptive and fixed codebook optimization schemes described in Section 10.5.2.4. These joint optimization schemes are described below.

The simplest optimization scheme - Method A from Section 10.5.2.4 - involves calculating the adaptive and fixed codebook gains and indices as usual, and then updating the two gains for the given codebook indices k and α using Equations 10.28 and 10.29, which are repeated here for convenience:

$$G_1 = \frac{C_\alpha \xi_k - C_k Y_{\alpha k}}{\xi_\alpha \xi_k - Y_{\alpha k}^2} \quad (12.53)$$

$$G_2 = \frac{C_k \xi_\alpha - C_\alpha Y_{\alpha k}}{\xi_\alpha \xi_k - Y_{\alpha k}^2}. \quad (12.54)$$

Here G_1 is the LTP gain, G_2 is the fixed codebook gain,

$$\xi_\alpha = \sum_{n=0}^{vs-1} y_\alpha^2(n) \quad (12.55)$$

is the energy of the filtered adaptive codebook signal and

$$C_\alpha = \sum_{n=0}^{vs-1} x(n) y_\alpha(n) \quad (12.56)$$

is the correlation between the filtered adaptive codebook signal and the codebook target $x(n)$. Similarly ξ_k is the energy of the filtered fixed codebook signal $[c_k(n) * h(n)]$, and C_k is the correlation between this and the target signal. Finally

$$Y_{\alpha k} = \sum_{n=0}^{vs-1} y_\alpha(n) [c_k(n) * h(n)] \quad (12.57)$$

is the correlation between the filtered signals from the two codebooks.

We studied the performance of this gain update scheme in our vector length 10 codec. A 7 bit fixed shape codebook was trained, but the LTP and fixed codebook gains were not quantized. We found that the gain update improved the segmental SNR of our codec by 1.2 dB over the codec with backward adapted 3 tap LTP and no fixed codebook gain quantization. This is a much more significant improvement than that reported in Section 10.5.2.4 for our 4.7 kbits/s ACELP codec, because of the much higher update rate for the gains used in our present codec. In our low delay codec the two gains are calculated for every 10 sample vector, whereas in the 4.7 kbits/s ACELP codec used in Section 10.5 the two gains are updated only every 60 sample sub-frame.

Encouraged by these results we also invoked the second sub-optimal joint codebook search procedure described in Section 10.5.2.4. In this search procedure the adaptive codebook delay α is determined first, by backward adaption in our present codec,

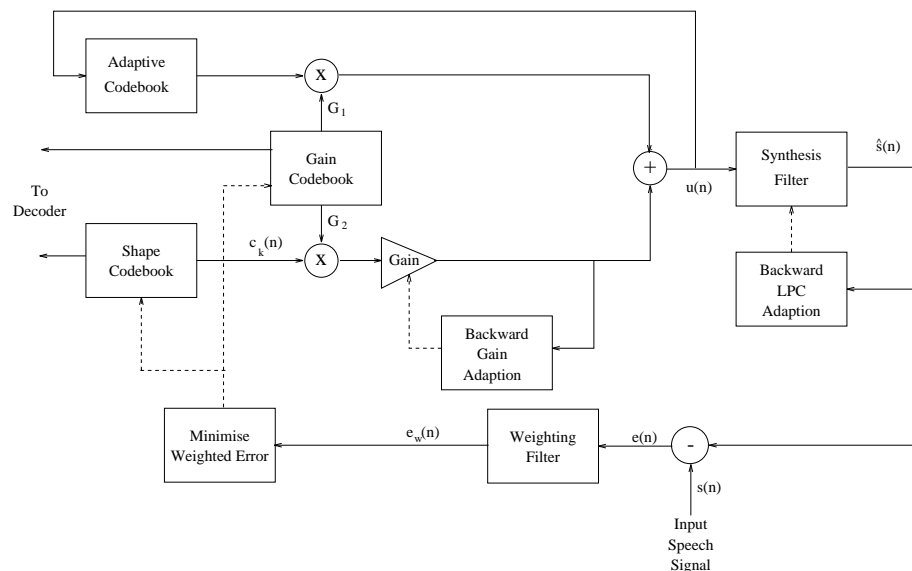


Figure 12.20: Scheme Three Low Delay CELP Codec

and then for each fixed codebook index k the optimum LTP and fixed codebook gains G_1 and G_2 are determined using Equations 12.53 and 12.54 above. The index k which maximises $T_{\alpha k}$, given below in Equation 12.58, will minimise the weighted error between the reconstructed and the original speech for the present vector, and is transmitted to the decoder. This codebook search procedure was referred to as Method B in Section 10.5.2.4.

$$T_{\alpha k} = 2(G_1 C_{\alpha} + \hat{\sigma} G_2 C_k - \hat{\sigma} G_1 G_2 Y_{\alpha k}) - G_1^2 \xi_{\alpha} - \hat{\sigma}^2 G_2^2 \xi_k \quad (12.58)$$

We trained a new 7-bit fixed shape codebook for this joint codebook search algorithm, and the two gains G_1 and G_2 were left unquantized. We found that this scheme gave an additional improvement in the performance of the codec so that its segmental SNR was now 2.7 dB higher than the codec with backward adapted 3 tap LTP and no fixed gain quantization. Again this is a much more significant improvement than that which we found for our 4.7 kbits/s ACELP codec.

12.9.4.2 Quantization of Jointly Optimized Gains

The improvements quoted above for our vector size 10 codec when we use an adaptive codebook arrangement with joint calculation of the LTP and fixed codebook gains, and no quantization of either gain, are quite promising. Next we considered the quantization of the two gains G_1 and G_2 . In order to minimise the number of bits used we decided to use a vector quantizer for the two gains. A block diagram of the coding scheme used is shown in Figure 12.20. We refer to this arrangement as “Scheme Three”.

This Scheme Three codec with forward adaptive LTP was tested with 4,5,6 and

7 bit vector quantizers for the fixed and adaptive codebook gains and a 7 bit shape codebook. The vector quantizers were trained as follows. For a given vector quantizer level i the total weighted energy E_i for speech vectors using this level will be

$$E_i = \sum_{m \in N_i} \left(\sum_{n=0}^{vs-1} (x_m(n) - G_{1i}y_{\alpha m}(n) - G_{2i}\hat{\sigma}_m[h_m(n) * c_m(n)])^2 \right). \quad (12.59)$$

Here $x_m(n)$, $y_{\alpha m}(n)$, and $h_m(n)$ are the signals $x(n)$, $y_{\alpha}(n)$, and $h(n)$ in the m 'th vector, $\hat{\sigma}_m$ is the value of the backward adapted gain $\hat{\sigma}$ in the m 'th vector, $c_m(n)$ is the fixed codebook entry $c_k(n)$ used in the m 'th vector, G_{1i} and G_{2i} are the values of the two gains in the i 'th entry of the joint vector quantizer, and N_i is the set of speech vectors that use the i 'th entry of the vector quantizer. As before vs is the vector size used in the codec, which in our present experiments is ten.

Expanding Equation 12.59 gives:

$$E_i = \sum_{m \in N_i} (X_m + G_{1i}^2 \xi_{\alpha m} + G_{2i}^2 \hat{\sigma}_m^2 \xi_{km} - 2G_{1i}C_{\alpha m} - 2\hat{\sigma}_m G_{2i}C_{km} + 2\hat{\sigma}_m G_{1i}G_{2i}Y_{\alpha km}) \quad (12.60)$$

where $X_m = \sum_{n=0}^{vs-1} x_m^2(n)$ is the energy of the target signal $x_m(n)$, and $\xi_{\alpha m}$, ξ_{km} , $C_{\alpha m}$, C_{km} and $Y_{\alpha km}$ are the values in the m 'th vector of ξ_{α} , ξ_k , C_{α} , C_k and $Y_{\alpha k}$ defined earlier.

Differentiating Equation 12.61 with respect to G_{1i} and setting the result to zero gives

$$\frac{\partial E_i}{\partial G_{1i}} = \sum_{m \in N_i} (2G_{1i}\xi_{\alpha m} - 2C_{\alpha m} + 2\hat{\sigma}_m G_{2i}Y_{\alpha km}) = 0 \quad (12.61)$$

or

$$G_{1i} \sum_{m \in N_i} \xi_{\alpha m} + G_{2i} \sum_{m \in N_i} \hat{\sigma}_m Y_{\alpha m} = \sum_{m \in N_i} C_{\alpha m}. \quad (12.62)$$

Similarly, differentiating with respect to G_{2i} and setting the result to zero gives:

$$G_{1i} \sum_{m \in N_i} \hat{\sigma}_m Y_{\alpha km} + G_{2i} \sum_{m \in N_i} \hat{\sigma}_m^2 \xi_{km} = \sum_{m \in N_i} \hat{\sigma}_m C_{km}. \quad (12.63)$$

Solving these two simultaneous equations gives the optimum values of G_{1i} and G_{2i} for the cluster of vectors N_i as:

$$G_{1i} = \frac{(\sum_{m \in N_i} C_{\alpha m})(\sum_{m \in N_i} \hat{\sigma}_m^2 \xi_{km}) - (\sum_{m \in N_i} \hat{\sigma}_m C_{km})(\sum_{m \in N_i} \hat{\sigma}_m Y_{\alpha km})}{(\sum_{m \in N_i} \xi_{\alpha m})(\sum_{m \in N_i} \hat{\sigma}_m^2 \xi_{km}) - (\sum_{m \in N_i} \hat{\sigma}_m Y_{\alpha km})^2} \quad (12.64)$$

and

$$G_{2i} = \frac{(\sum_{m \in N_i} \hat{\sigma}_m C_{km})(\sum_{m \in N_i} \xi_{\alpha m}) - (\sum_{m \in N_i} C_{\alpha m})(\sum_{m \in N_i} \hat{\sigma}_m Y_{\alpha km})}{(\sum_{m \in N_i} \xi_{\alpha m})(\sum_{m \in N_i} \hat{\sigma}_m^2 \xi_{km}) - (\sum_{m \in N_i} \hat{\sigma}_m Y_{\alpha km})^2}. \quad (12.65)$$

Using Equations 12.64 and 12.65 we performed a closed-loop training of the vector

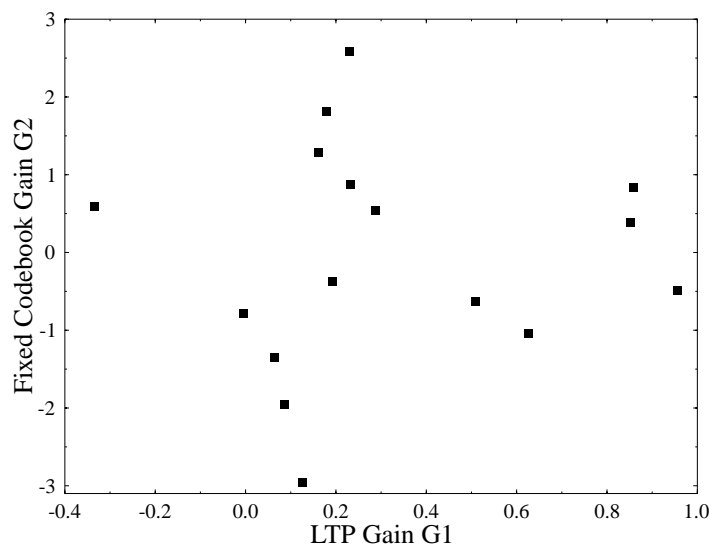


Figure 12.21: Values of G_1 and G_2 in the 4 Bit Gain Quantizer

quantizer gain codebook along with the fixed shape codebook similarly to the training of the shape and single gain codebooks described in Section 12.7. However we found a similar problem to that which we encountered when training scalar codebooks for G_1 and G_2 in Section 10.5.2.5. Specifically although almost all values of G_1 have magnitudes less than 2, a few values have very high magnitudes. This leads to a few levels in the trained vector quantizers having very high values, and being very rarely used. Following an in-depth investigation into this phenomenon we solved the problem by excluding all vectors for which the magnitude of G_1 was greater than 2, or the magnitude of G_2 was greater than 5, from the training sequence. This approach solved the problems of the trained gain codebooks having some very high and very rarely used levels.

We trained vector quantizers for the two gains using 4, 5, 6 and 7 bits. The values of the 4 bit trained vector quantizer for G_1 and G_2 are shown in Figure 12.21. It can be seen that when G_1 is close to zero, the values of G_2 have a wide range of values between -3 and +3, but when the speech is voiced and G_1 is high the fixed codebook contribution to the excitation is less significant, and the quantized values of G_2 are closer to zero.

Our trained joint gain codebooks are searched as follows. For each fixed codebook entry k the optimum gain codebook entry is found by tentatively invoking each pair of gain values in Equation 12.58, in order to test which level maximises $T_{\alpha k}$ and hence minimises the weighted error energy. The segmental SNR of our Scheme Three

Gain Codebook Bits	Segmental SNR (dB)
4 Bits	14.81
5 Bits	15.71
6 Bits	16.54
7 Bits	17.08

Table 12.13: Performance of the Scheme Three Codecs

codec with a trained 7 bit shape codebook and trained 4,5,6 and 7 bit joint G_1/G_2 vector quantizers is shown in Table 12.13. The segmental SNRs in this table should be compared with the value of 14.29 dB obtained for the Scheme One codec with a 3 bit scalar quantizer for G_2 and 3 tap backward adapted LTP.

It can be seen from Table 12.13 that the joint G_1/G_2 gain codebooks give a steady increase in the performance of the codec as the size of the gain codebook is increased. In the next Section we describe the use of backward adaptive voiced/unvoiced switched codebooks to further improve the performance of our codec.

12.9.4.3 8-4kbps Codecs - Voiced/Unvoiced Codebooks

In Section 12.7 we discussed using different codebooks for voiced and unvoiced segments of speech, and using a backward adaptive voicing decision to select which codebooks to use. However we found that in the case of a codec with fully backward adaptive LTP no significant improvement in the codec's performance was achieved by using switched codebook excitation. In this Section we discuss using a similar switching arrangement in conjunction with our Scheme Three codec described above.

The backward adaptive voiced/unvoiced switching is based on the voiced/unvoiced switching used in the postfilter employed in the G728 codec [231]. In our codec the switch uses the normalised autocorrelation value of the past reconstructed speech signal $\hat{s}(n)$ at the delay α which is used by the adaptive codebook. This normalised autocorrelation value β_α is given by

$$\beta_\alpha = \frac{\sum_{n=-100}^{-1} \hat{s}(n)\hat{s}(n-\alpha)}{\sum_{n=-100}^{-1} \hat{s}^2(n-\alpha)}, \quad (12.66)$$

and when it is greater than a set threshold the speech is classified as voiced; otherwise the speech is classified as unvoiced. In our codec, as in the G728 postfilter, the threshold is set to 0.6.

Figure 12.22 shows a segment of the original speech and the normalised autocorrelation value β_α calculated from the reconstructed speech of our 8 kbits/s codec. To aid the clarity of this graph the values of β_α have been limited to lie between 0.05 and 0.95. It can be seen that the condition $\beta_\alpha > 0.6$ gives a good indication of whether the speech is voiced or unvoiced.

The backward adaptive voicing decision described above was incorporated into our Scheme Three codec shown in Figure 12.20 to produce a new coding arrangement which we referred to as "Scheme Four". Shape and joint gain codebooks were trained as described earlier for both the voiced and unvoiced modes of operation in a vector

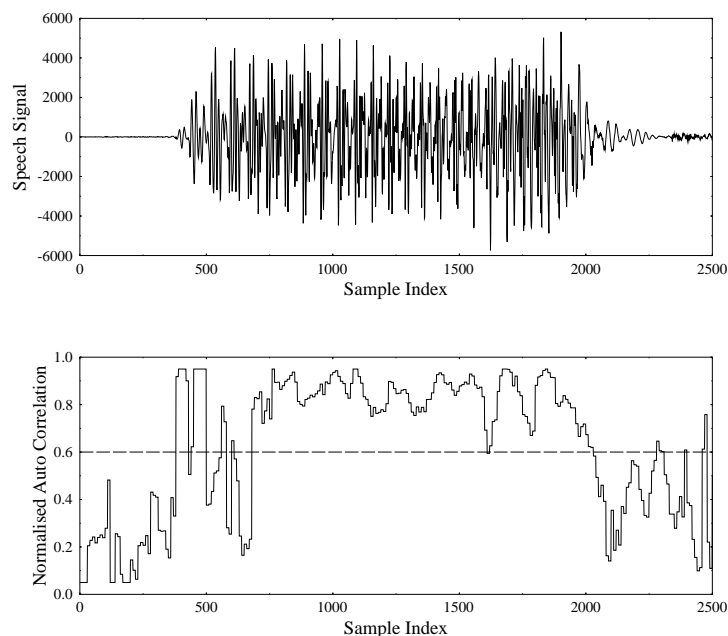


Figure 12.22: Normalised Autocorrelation Value β_α During Voiced and Unvoiced Speech

Gain Codebook Bits	Segmental SNR (dB)
4 Bits	15.03
5 Bits	15.92
6 Bits	16.56
7 Bits	17.12

Table 12.14: Performance of the Scheme Four Codecs

length 10 codec. The quantized values of G_1 and G_2 in both the 4 bit voiced and unvoiced codebooks are shown in Figure 12.23. It can be seen that similarly to Figure 12.21 when G_1 is high the range of values of G_2 is more limited than when G_1 is close to zero. Furthermore, as expected, the voiced codebook has a group of quantizer levels with G_1 close to one, whereas the values of the LTP gain in the unvoiced codebook are closer to zero.

The results we achieved with seven bit shape codebooks and joint gain codebooks of various sizes are shown in Table 12.14. It can be seen by comparing this to Table 12.13 that the voiced/unvoiced switching gives an improvement in the codec's performance of about 0.25 dB for the 4 and the 5 bit gain quantizers, and a smaller improvement for the 6 and 7 bit gain quantizers.

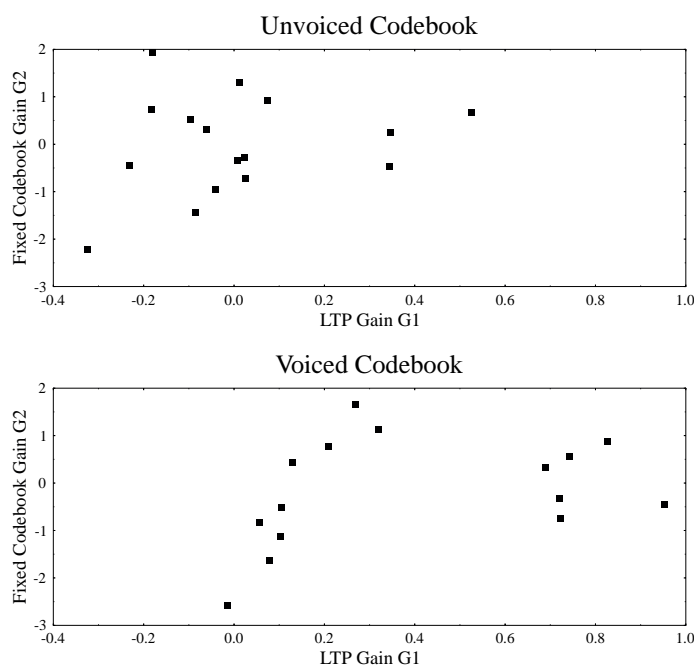


Figure 12.23: Values of G_1 and G_2 in the 4 Bit Voiced and Unvoiced Gain Quantizers

12.9.5 Low Delay Codecs at 4-8 kbits/s

In the previous three Sections we have considered the improvements that can be achieved in our vector size 10 codec by increasing the size of the shape and gain codebooks, and by using forward adaption of the short term predictor coefficients and the long term predictor gain. The improvements obtained by these schemes are summarised in Table 12.15, which shows the various gains in the codec's segmental SNR against the number of extra bits used to represent each ten sample vector.

In this table the Scheme One codec (see Section 12.9.2) is the vector size 10 codec, with three-tap backward adapted LTP and a 20-tap backward adapted short term predictor. The table shows the gains in the segmental SNR of the codec that are achieved by adding one or two extra bits to the shape or the scalar gain codebooks.

The Scheme Two codec (see Section 12.9.3) also uses 3 tap backward adapted LTP, but uses forward adaption to determine the short term synthesis filter coefficients. Using these coefficients without quantization gives an improvement in the codecs segmental SNR of 0.82 dB, which would be reduced if quantization were applied. In reference [143], where forward adaption is used for the LPC parameters, 19 bits are used to quantize a set of LSFs for every 80 sample frame; this quantization scheme would require us to use about 2.4 extra bits per 10 sample vector.

The Scheme Three codec (see Section 12.9.4) uses backward adaption to determine the short term predictor coefficients and the long term predictor delay. However

	Synthesis Filter	Long Term Predictor	Shape C.B.	Gain C.B.	Extra Bits	Δ Seg. SNR
Scheme One	Backward Adapted $p=20$	3 Tap Backward Adapted	7 Bits	3 Bits	0	0 dB
			7 Bits	4 Bits	1	+0.95 dB
			8 Bits	3 Bits	1	+1.04 dB
			7 Bits	5 Bits	2	+1.33 dB
			8 Bits	4 Bits	2	+1.72 dB
			9 Bits	3 Bits	2	+1.83 dB
Scheme Two	Forward Adapted $p=10$	3 Tap Backward Adapted	7 Bits	3 Bits	≈ 2.4	$\leq +0.82$ dB
Scheme Three	Backward Adapted $p=20$	Forward Adapted	7 Bits	4 Bits	1	+0.52 dB
			7 Bits	5 Bits	2	+1.42 dB
			7 Bits	6 Bits	3	+2.25 dB
			7 Bits	7 Bits	4	+2.79 dB
Scheme Four	Backward Adapted $p=20$	Switched	7 Bits	4 Bits	1	+0.74 dB
		Forward	7 Bits	5 Bits	2	+1.63 dB
		Adapted	7 Bits	6 Bits	3	+2.27 dB
			7 Bits	7 Bits	4	+2.83 dB

Table 12.15: Improvements Obtained Using Schemes One to Four

forward adaption is used to find the LTP gain, which is jointly determined along with the fixed codebook index and gain. The LTP gain and the fixed codebook gain are jointly vector quantized using 4, 5, 6 or 7 bit quantizers, which implies using between 1 and 4 extra bits per 10 sample vector.

Finally the Scheme Four codec uses the same coding strategy as the Scheme Three codec, but also implements a backward adapted switch between specially trained shape and vector gain codebooks for the voiced and unvoiced segments of speech.

It is clear from Table 12.15 that, for our vector size 10 codec, using extra bits to allow forward adaption of the synthesis filter parameters is the least efficient way of using these extra bits. If we were to use two extra bits the largest gain in the codec's segmental SNR is given if we simply use the Scheme One codec and increase the size of the shape codebook by 2 bits. This gain is almost matched if we allocate one extra bit to both the shape and gain codebooks in the Scheme One codec, and this would increase the codebook search complexity less dramatically than allocating both extra bits to the shape codebook.

In order to give a fair comparison between the different coding schemes at bit rates between 4 and 8 kbits/s we tested the Scheme One, Scheme Three and Scheme Four codecs using 8 bit shape codebooks, 4 bit gain codebooks and vector sizes of 12, 15, 18 and 24 samples. This gave three different codecs at 8, 6.4, 5.3 and 4 kbits/s. Note that as the vector size of the codecs increase, their complexity also increases. Methods of reducing this complexity are possible [349], but have not been studied in our work. The segmental SNRs of our three 4-8 kbits/s codecs against their bit rates is shown in Figure 12.24.

Several observations can be made from this graph. At 8 kbits/s, as expected from the results in Table 12.15, the Scheme One codec gives the best quality reconstructed

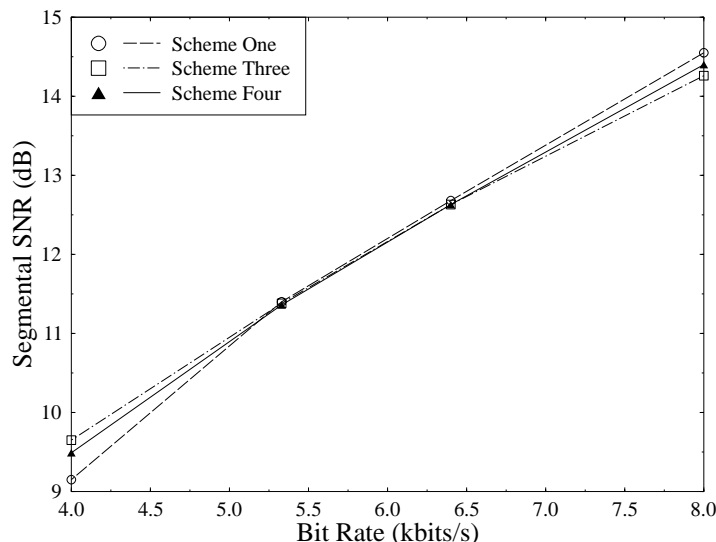


Figure 12.24: Performance of Schemes One Three and Four Codecs at 4-8 kbits/s

speech, with a segmental SNR of 14.55 dB. However as the vector size is increased and hence the bit rate reduced it is the Scheme One codec whose performance is most badly affected. At 6.4 kbits/s and 5.3 kbits/s all three codecs give very similar segmental SNRs, but at 4 kbits/s the Scheme One codec is clearly worse than the other codecs, which use forward adaption of the LTP gain. This indicates that although the three tap backward adapted LTP is very effective at 8 kbits/s and above, it is less effective as the bit rate is reduced. Furthermore the backward adaptive LTP scheme is more prone to channel error propagation.

Similarly, as indicated in Table 12.15, the backward adaptive switching between specially trained voiced and unvoiced gain and shape codebooks improves the performance of our Scheme Four codec at 8 kbits/s so that it gives a higher segmental SNR than the Scheme Three codec. However as the bit rate is reduced the gain due to this codebook switching is eroded, and at 4 kbits/s the Scheme Four codec gives a lower segmental SNR than the Scheme Three codec. This is due to inaccuracies in the backward adaptive voicing decisions at the lower bit rates. Figure 12.25 shows the same segment of speech as was shown in Figure 12.22, and the normalised autocorrelation value β_α calculated from the reconstructed speech of our Scheme Four codec at 4 kbits/s. It can be seen that the condition $\beta_\alpha > 0.6$ no longer gives a good indication of the voicing of the speech. Again for clarity of display the values of β_α have been limited to between 0.05 and 0.95 in this figure.

In listening tests we found that all three codecs gave near toll quality speech at 8

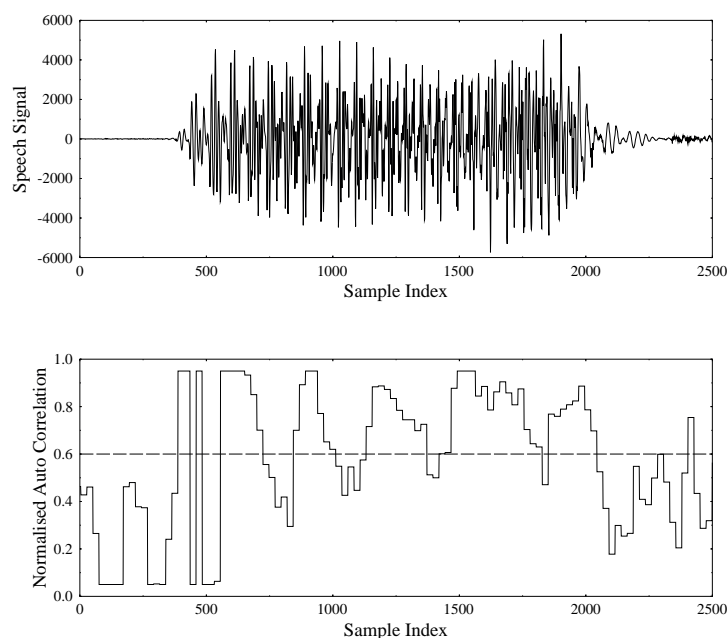


Figure 12.25: Normalised Autocorrelation Value β_α During Voiced and Unvoiced Speech

kbits/s, with differences between the codecs being difficult to distinguish. However, at 4 kbits/s the Scheme Two codec sounded clearly better than the Scheme One codec, and gave reconstructed speech of communications quality.

12.9.6 Low Delay ACELP Codec

In this Section of our work on low delay CELP codecs operating between 4 and 8 kbits/s we implemented a low delay version of our Algebraic CELP (ACELP) codec which was described in Section 10.4.3. We developed a series of low delay codecs with a frame size of 40 samples or 5 ms, and hence a total delay of about 15 ms, and with various bit rates between 5 and 6.2 kbits/s. All of these codecs use backward adaption with the recursive windowing function described in Section 12.4.2 in order to determine the coefficients for the synthesis filter, which has an order of $p = 20$. Furthermore, they employ the same weighting filter, which was described in Section 12.4.1, as our other low delay codecs. However apart from this they have a structure similar to the codecs described in Section 10.4.3. An adaptive codebook is used to represent the long term periodicities of the speech, with possible delays taking all integer values between 20 and 147 and being represented using 7 bits. As described in Section 10.4.3 the best delay is calculated once per 40 sample vector within the Analysis-by-Synthesis loop at the encoder, and then transmitted to the decoder.

Initially we used the 12 bit ACELP fixed codebook structure shown in Table 10.4

Pulse Number i	Amplitude	Possible Position m_i
0	+1	1,6,11,16,21,26,31,36
1	-1	2,7,12,17,22,27,32,37
2	+1	3,8,13,18,23,28,33,38
3	-1	4,9,14,19,24,29,34,39

Table 12.16: Pulse Amplitudes and Positions for the 12 bit ACELP Codebook

which is repeated here in Table 12.16. Each 40 sample vector has a fixed codebook signal given by 4 non-zero pulses of amplitude +1 or -1, whose possible positions are shown in Table 12.16. Each pulse position is encoded with 3 bits giving a 12 bit codebook. As it was explained in Section 10.3, the pulse positions can be found using a series of four nested loops, leading to a very efficient codebook search algorithm [178, 280].

In our first low delay ACELP codec, which we refer to as Codec A, we used the same 3 and 5 bit scalar quantizers as were used in the codecs in Section 10.4.3 to quantize the adaptive and fixed codebook gains G_1 and G_2 . This meant that 12 bits were required to represent the fixed codebook index, 7 bits for the adaptive codebook index and a total of 8 bits to quantize the two codebook gains. This gave a total of 27 bits to represent each 40 sample vector, giving a bit rate for this codec of 5.4 kbits/s. We found that this codec gave an average segmental SNR of 10.20 dB, which should be compared to the average segmental SNRs for the same speech files of 9.83 dB, 11.13 dB and 11.42 dB for our 4.7 kbits/s, 6.5 kbits/s and 7.1 kbits/s forward adaptive ACELP codecs described in Section 10.4.3. All of these codecs have a similar level of complexity, but the backward adaptive 5.4 kbits/s ACELP codec has a frame size of only 5 ms, compared to the frame sizes of 20 or 30 ms for the forward adaptive systems. Furthermore it can be seen from Figure 12.26 that, upon interpolating the segmental SNRs between the three forward adaptive ACELP codecs, the backward adaptive ACELP codec at 5.4 kbits/s gives a very similar level of performance to the forward adaptive codecs. In this figure we have marked the segmental SNRs of the three forward adaptive ACELP codecs with circles, and the segmental SNR of our low delay ACELP codec at 5.4 kbits/s with a diamond. Also marked with diamonds are the segmental SNRs and bit rates of other backward adaptive ACELP codecs which will be described later. For comparison the performance of the Scheme One low delay codec, described in Section 12.9.5 and copied from Figure 12.24, is also shown.

It can be seen from Figure 12.26 that although the 5.4 kbits/s low delay backward adaptive ACELP codec described above gives a similar performance in terms of segmental SNR to the higher delay forward adaptive ACELP codecs, it performs significantly worse than the Scheme One codec of Table 12.15, which uses a shorter vector size and a trained shape codebook. We therefore attempted to improve the performance of our low delay ACELP codec by introducing vector quantization and joint determination of the two codebook gains G_1 and G_2 . Note that similar vector quantization and joint determination of these gains was used in the Scheme Three and Scheme Four codecs described in Section 12.9.5. We also re-introduced the backward adaption of the fixed codebook gain G_2 - as known from the schematic of the G.728

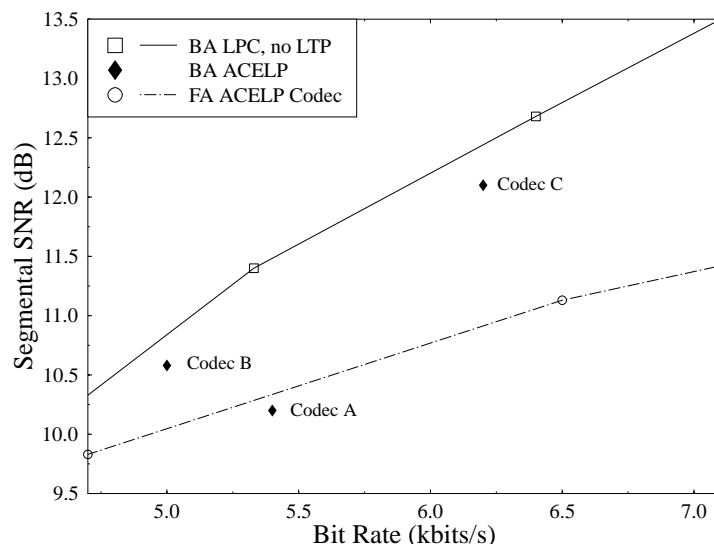


Figure 12.26: Performance of Low Delay ACELP codecs

decoder seen in Figure 12.2, which was used in our other low delay codecs as detailed in Section 12.4.3. We replaced the 3- and 5-bit scalar quantizers for G_1 and G_2 with a 6-bit joint vector quantizer for these gains, which resulted in a total of 25 bits being used to represent each 40 sample vector and therefore gave us a 5 kbits/s codec. We refer to this as Codec B. The joint 6-bit vector quantizer for the gains was trained as described in Section 12.9.4.2. A joint codebook search procedure was used so that for each fixed codebook index k the joint gain codebook was searched to find the gain codebook index which minimised the weighted error for that fixed codebook index. The best shape and gain codebook indices are therefore determined together. This codebook search procedure results in a large increase in the complexity of the codec, but also significantly increases the performance of the codec.

We found that our 5 kbits/s Codec B, using joint vector quantization of G_1 and G_2 and backward adaption of G_2 , gave an average segmental SNR of 10.58 dB. This is higher than the segmental SNR of the codec with scalar gain quantization, ie Codec A, despite Codec B having a lower bit rate. The performance of this Codec B is marked with a diamond in Figure 12.26, which shows that it falls between the segmental SNRs of the ACELP codecs with scalar gain quantization and the Scheme One codecs.

Next we replaced the 12 bit algebraic codebook detailed in Table 12.16 with the 17 bit algebraic codebook used in the G.729 ACELP codec described in Section 11.8. Also the 6 bit vector quantization of the two gains was replaced with 7 bit vector quantization. This gave a 6.2 kbits/s codec, referred to as Codec C, which is similar

	Algebraic Codebook	Gain Quantization	Bit Rate (kbits/s)	Segmental SNR
Codec A	12 Bit	3+5 Bit Scalar	5.4	10.2 dB
Codec B	12 Bit	6 Bit Vector	5	10.6 dB
Codec C	17 Bit	7 Bit Vector	6.2	12.1 dB

Table 12.17: Performance and Structure of Low Delay ACELP Codecs

to the G.729 codec. The main difference between G.729 and our Codec C is that G.729 uses forward adaption to determine the LPC coefficients, whereas Codec B uses backward adaption. This implies that it does not transmit the 18 bits per 10ms that G.729 uses to represent the LPC parameters, and hence it operates at a bit rate 1.8 kbits/s lower. Also its buffering delay is halved to only 5 ms.

We found that this Codec C gave reconstructed speech with a segmental SNR of 12.1 dB, as shown in Figure 12.26. It can be seen that our G.729 like codec gives a better segmental SNR than the forward adaptive ACELP codecs described earlier. This is because of the more advanced 17 bit codebook, together with the joint determination and vector quantization of the fixed and the adaptive codebook gains, used in the backward adaptive ACELP codec. It is also clear from Figure 12.26 that Codec C gives a similar performance to the backward adaptive variable rate codecs with trained codebooks. Subjectively we found that Codec C gave speech of good communications quality, but significantly lower than that of the toll quality produced by the forward adaptive G729. Even so this codec may be preferred to G.729 in situations where a lower bit rate and delay are required, and the lower speech quality can be accepted.

The characteristics of our low delay ACELP codecs are summarised in Table 12.17. In the next Section we discuss error sensitivity issues relating to the low delay codecs described in this chapter.

12.10 Backward-adaptive Error Sensitivity Issues

Traditionally one serious disadvantage of using backward adaption of the synthesis filter is that it is more sensitive to channels errors than forward adaption. In this Section we first consider the error sensitivity of the 16 kbits/s G728 codec described earlier. We then discuss the error sensitivity of the 4-8 kbits/s low delay codecs described earlier, and means of improving this error sensitivity. Finally we investigate the error sensitivity of our low delay ACELP codec described above, and compare this to the error sensitivity of a traditional forward adaptive ACELP codec.

12.10.1 The Error Sensitivity of the G728 Codec

As described earlier for each five sample speech vector the G728 codec produces a 3 bit gain codebook index, and an eight bit shape codebook index. Figure 12.27 shows the sensitivity to channel errors of these ten bits. The error sensitivities were measured by, for each bit, corrupting the given bit only with a 10% Bit Error Rate (BER). This approach was taken, rather than the more usual method of corrupting the given bit in

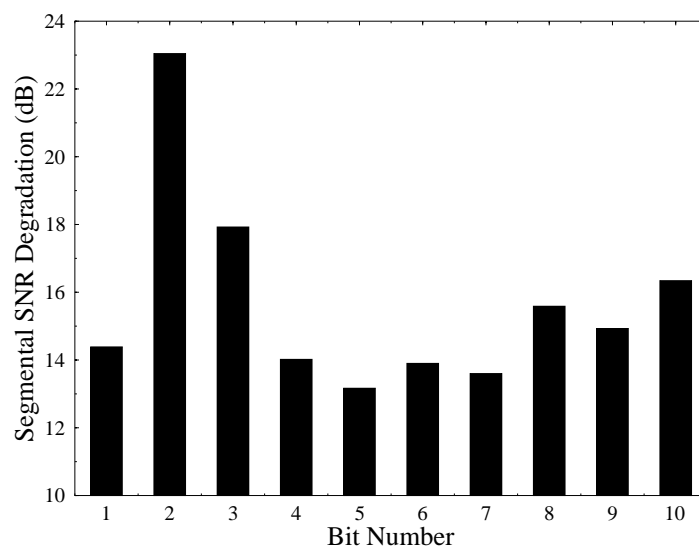


Figure 12.27: Degradation in G728 Segmental SNR Caused by 10 % BER in Given Bits

every frame, to allow account to be taken of the possible different error propagation properties of different bits [52]. Bits 1 and 2 in Figure 12.27 represent the magnitude of the excitation gain, bit 3 represents the sign of this gain, and the remaining bits are used to code the index of the shape codebook entry chosen to represent the excitation. It can be seen from this figure that not all ten bits are equally sensitive to channel errors. Notice for example that bit 2, representing the most significant bit of the excitation gain's magnitude, is particularly sensitive.

This unequal error sensitivity can also be seen from Figure 12.28, which shows the segmental SNR of the G728 codec for channel BERs between 0.001% and 1%. The solid line shows the performance of the codec when the errors are equally distributed amongst all ten bits, whereas the dashed lines show the performance when the errors are confined only to the 5 most sensitive bits (the so called "Class One" bits) or the 5 least sensitive bits (the "Class Two" bits). The ten bits were arranged into these two groups based on the results shown in Figure 12.27 – bits 2,3,8,9 and 10 formed Class One and the other five bits formed Class Two. It can be seen that the Class One bits are about two or three times more sensitive than the Class Two bits. Therefore it is clear that when the G728 codec is employed in an error-prone transmission scheme, for example in a mobile radio transmission system, the error resilience of the system will be improved if un-equal error protection is employed [110]. The use of un-equal error protection for speech codecs is discussed in detail later.

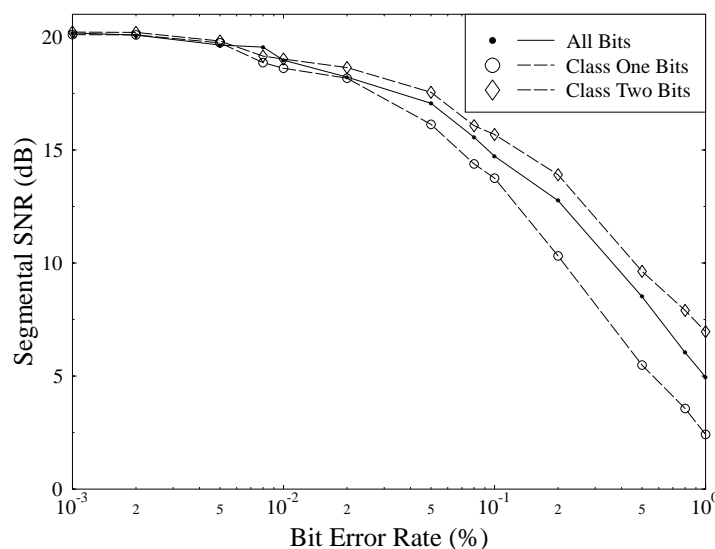


Figure 12.28: Segmental SNR of G728 Codec Against Channel BER

12.10.2 The Error Sensitivity of Our 4-8 kbits/s Low Delay Codecs

We now consider the error sensitivity of some of our 4-8 kbits/s codecs which were described in Section 12.9.5. It is well known that codecs using backward adaption for both the LTP delay and gain are very sensitive to bit errors, and this is why LTP was not used in G728 [217]. Thus, as expected, we found that the Scheme One codec gave a very poor performance, when subjected to even a relatively low Bit Error Rate (BER). Unfortunately, we also found similar results for the Scheme Three and Scheme Four codecs, which, although they used backward adaption for the LTP delay, used forward adaption for the LTP gain. We therefore decided that none of these codecs are suitable for use over error-prone channels. However the Scheme One codec can be easily modified by removing its entirely backward adapted 3 tap LTP, and increasing the order of its short term filter to 50 as in G728, to make it less sensitive to channel errors. Although this impairs the performance of the codec, as can be seen from Figure 12.29 the resulting degradation in the codec's segmental SNR is not too serious, especially at low bit rates. Therefore in this Section we detail the error sensitivity of the Scheme One codec with its LTP removed, and describe means of making this codec less sensitive to channel errors. For simplicity, only the error sensitivity of the codec operating with a frame-length of 15 samples and a bit rate of 6.4 kbits/s are detailed in this Section. However similar results also apply at the

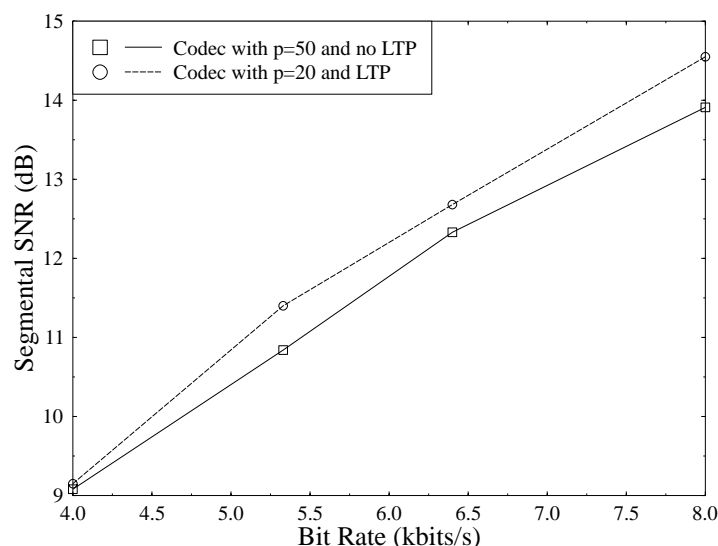


Figure 12.29: Segmental SNR of G728 Codec Against Channel BER

other bit rates.

At 6.4 kbits/s our codec transmits only 12 bits per 15 sample frame from the encoder to the decoder. Of these 12 bits 8 are used to represent the index of the shape codebook, and the remaining 4 bits are used to represent the index of the gain codebook entry used. The error resilience of these bits can be significantly improved by careful assignment of codebook indices to the various codebook entries. Ideally, each codebook entry would be assigned an index so that corruption of any of the bits representing this index will result in another entry being selected in the decoder's codebook which is in some way "close" to the intended codebook entry. If this ideal can be achieved, then the effects of errors in the bits representing the codebook indices will be minimised.

Consider first the 8 bit shape codebook. Initially the 256 available codebook indices are effectively randomly distributed amongst the codebook entries. We seek to rearrange these codebook indices so that when the index representing a codebook entry is corrupted, the new index will represent a codebook entry that is "close" to the original entry. In our work we chose to measure this "closeness" by the squared error between the original and the corrupted codebook entries. We considered only the effects of single bit errors among the 8 codebook bits because at reasonable Bit Error Rates (BERs) the probability of two or more errors occurring in 8 bits will be small. Thus for each codebook entry the "closeness" produced by a certain arrangement of codebook entries is given by the sum of the squared errors between the original codebook

entry and the eight corrupted entries that would be produced by inverting each of the 8 bits representing the entry's index. The overall "cost" of a given arrangement of codebook indices is then given by the closeness for each codebook entry, weighted by the probability of that codebook entry being used. Thus the cost we seek to minimise is given by

$$\text{Cost} = \sum_{j=0}^{255} P(j) \left[\sum_{i=1}^8 \left(\sum_{n=1}^{15} (c_j(n) - c_j^i(n))^2 \right) \right] \quad (12.67)$$

where $P(j)$ is the probability of the j 'th codebook entry being used, $c_j(n)$, $n = 1 \dots 15$, is the j 'th codebook entry and $c_j^i(n)$ is the entry that will be received if the index j is transmitted but the i 'th bit of this index is corrupted.

The problem of choosing the best arrangement of the 256 codebook indices among the codebook entries is similar to the famous travelling salesman problem. In this problem the salesman must visit each of N cities, and must choose the order in which he visits the cities so as to minimise the total distance he travels. As N becomes large it becomes impractical to solve this problem using an exhaustive search of all possible orders in which he could visit the cities - the complexity of such a search is proportional to $N!$ Instead a non-exhaustive search must be used which we hope will find the best order possible in which to visit the N cities.

The minimisation method of simulated annealing has been successfully applied to this problem [111], and has also been used by other researchers as a method of improving the error resilience of quantizers [350]. Simulated annealing works, as its name suggests, in analogy to the annealing (or slow cooling) of metals. When metals cool slowly from their liquid state they start in a very disordered and high energy state and reach equilibrium in an extremely ordered crystalline state. This crystal is the minimum energy state for the system, and simulated annealing similarly allows us to find the global minimum of a complex function with many local minima. The procedure works as follows. The system starts in an initial state, which in our situation is an initial assignment of the 256 codebook indices to the codebook entries. A temperature like variable T is defined, and possible changes to the state of the system are randomly generated. For each possible change the difference ΔCost in the cost between the present state and the possible new state is evaluated. If this is negative, ie the new state has a lower cost than the old state, then the system always moves to the new state. If on the other hand ΔCost is positive then the new state has a higher cost than the old state, but the system may still change to this new state. The probability of this happening is given by the Boltzmann distribution

$$\text{prob} = \exp \left(\frac{-\Delta\text{Cost}}{kT} \right) \quad (12.68)$$

where k is a constant. The initial temperature is set so that kT is much larger than any ΔCost that is likely to be encountered, so that initially most offered moves will be taken. As the optimization proceeds the 'temperature' T is slowly decreased, and the number of moves to states with higher costs reduces. Eventually kT becomes so small that no moves with positive ΔCost are taken, and the system comes to equilibrium in what is hopefully the global minimum of its cost.

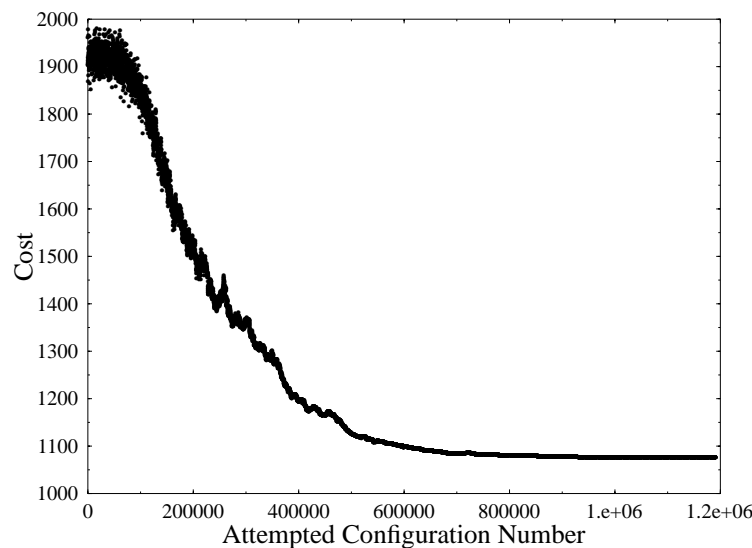


Figure 12.30: Reduction in Cost Using Simulated Annealing

The advantage of simulated annealing over other optimization methods is that it should not be deceived by local minima and should slowly make its way towards the global minimum of the function to be minimised. In order to make this likely to happen it is important to ensure that the temperature T starts at a high enough value, and is reduced suitably slowly. We followed the suggestions in [111] and reduced T by 10% after every $100N$ offered moves, or every $10N$ accepted moves, where N is the number of codebook entries (256). The initial temperature was set so that kT was equal to ten times the highest value of $\Delta Cost$ that was initially encountered. The random changes in the state of the system were generated by randomly choosing two codebook entries and swapping the indices of these two entries.

The effectiveness of the simulated annealing method in reducing the cost given in Equation 12.67 is shown in Figure 12.30. This graph shows the cost of the present arrangement of codebook indices against the number of arrangements of codebook indices which have been tried by the minimisation process. The initial randomly assigned arrangement of indices to codebook entries gives a cost of 1915. As can be seen in Figure 12.30 initially the temperature T is high and so many index assignments which have a higher cost than this are accepted. However slowly as the number of attempted configurations increases the temperature T decreases, and so fewer re-arrangements which increase the cost of the present arrangement are accepted. Thus as can be seen in Figure 12.30 the cost of present arrangement slowly falls, and the curve narrows as the temperature increases and less re-arrangements which increase

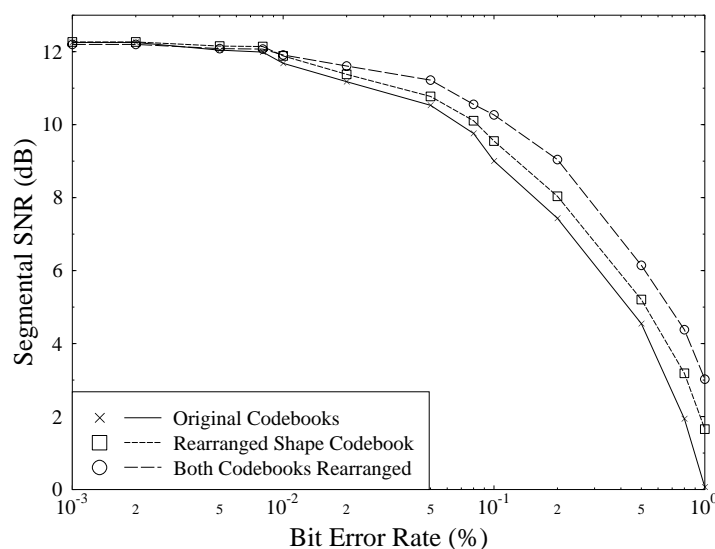


Figure 12.31: The Error Sensitivity of Our Low Delay 6.4 kbits/s Codec

the cost of the present arrangement are accepted. The cost of the final arrangement of codebook indices to codebook entries is 1077, which corresponds to a reduction in the cost of about 44%.

The effectiveness of this re-arrangement of codebook indices in increasing the resilience of the codec to errors in the bit stream between its encoder and decoder can be seen in Figure 12.31. This graph shows the variation in the segmental SNR of our 6.4 kbits/s low delay codec with the Bit Error Rate (BER) between its encoder and decoder. The solid line shows the performance of the codec with the original codebook index assignment, and the lower dashed line shows the performance when the shape codebook indices are re-arranged as described above. It can be seen that at BERs of between 0.1% and 1% the codec with the re-arranged codebook indices has a segmental SNR about 0.5 to 1 dB higher than the original codec.

Apart from the 8 shape codebook bits which the codec transmits from its encoder to the decoder, the only other information that is explicitly transmitted are the 4 bits representing the gain codebook entry selected. Initially indices were assigned to the 16 gain codebook entries using the simple Natural Binary Code (NBC). However because the gain codebook levels do not have an equiprobable distribution this simple assignment can be improved upon in a similar way to that described for the shape codebook described above. Again we defined a cost function that was to be minimised. This cost function was similar to that given in Equation 12.67 except because the gain codebook is scalar, whereas the shape codebook has a vector dimension of 15,

no summation over n is needed in the cost function for the gain codebook index arrangement. We used simulated annealing again to reduce the cost function over that given using a NBC, and found that we were able to reduce the cost by over 60%. The effect of this re-arrangement of the gain codebook indices is shown by the upper curve in Figure 12.31 which gives the performance of the Scheme One codec, with LTP removed, with both the gain and shape codebooks re-arranged. It can be seen that the re-arrangement of the gain codebook indices gives a further improvement in the error resilience of the codec, and that the codec with both the shape and gain codebooks re-arranged has a segmental SNR more than 1 dB higher than the original codec at BERs around 0.1%.

12.10.3 The Error Sensitivity of Our Low Delay ACELP Codec

The segmental SNR of our 6.2 kbits/s low delay ACELP codec described in Section 12.9.6 is shown in Figure 12.32. Also shown in this figure are the error sensitivities of our 6.4 kbits/s Scheme One codec with no LTP, and of a traditional 6.5 kbits/s forward adaptive ACELP codec. As noted above, at 0% BER the two backward adaptive codecs give similar segmental SNRs, but the forward adaptive codec gives a segmental SNR of about 1 dB lower. However in subjective listening tests the better spectral match provided by the forward adaptive codec, which is not adequately reflected in the segmental SNR distortion measure, results in it providing better speech quality than the two backward adaptive codecs. As the BER is increased the backward adaptive ACELP is the worst affected, but surprisingly, the other backward adaptive codec is almost as robust to channel errors as the forward adaptive ACELP codec. Both these codecs give a graceful degradation in their reconstructed speech quality at BERs up to about 0.1%, but provide impaired reconstructed speech for BERs much above this.

Let us now in the next Section provide an application scenario for employing the previously designed G.728-like 8-16 kbps speech codecs and evaluate the performance of the transceiver proposed.

12.11 A Low-Delay Multimode Speech Transceiver

12.11.1 Background

The intelligent, adaptively reconfigurable wireless systems of the near future require programmable source codecs in order to optimally configure the transceiver to adapt to time-variant channel and traffic conditions. Hence we designed a flexible transceiver for the previously portrayed programmable 8-16 kbits/s low-delay speech codec, which is compatible with the G.728 16 kbits/s ITU codec at its top rate and offers a graceful trade-off between speech quality and bit rate in the range 8-16 kbits/s. Source-matched Bose-Chaudhuri-Hocquenghem (BCH) codecs combined with un-equal protection pilot-assisted 4- and 16-level quadrature amplitude modulation (4-QAM, 16-QAM) are employed in order to transmit both the 8 and the 16 kbits/s coded speech bits at a signalling rate of 10.4 k Bd. In a bandwidth of 1728 kHz, which is used by the Digital European Cordless Telephone (DECT) system 55 duplex or 110 simplex time slots can be created. We will show that good toll quality speech is delivered in

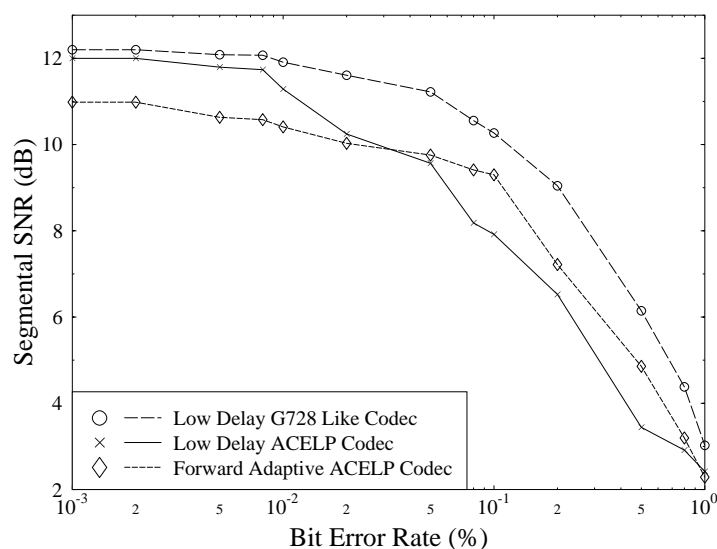


Figure 12.32: A Comparison of the Bit Error Sensitivities of Backward and Forward Adaptive Codecs

an equivalent user bandwidth of 15.71 kHz, if the channel signal-to-noise ratio (SNR) and signal-to-interference ratio (SIR) are in excess of about 18 and 26 dB for the lower and higher speech quality 4-QAM and 16-QAM modes, respectively.

12.11.2 8-16 kbps Codec Performance

The segmental SNR versus bit rate performance of our 8-16 kbits/s codec was shown in Figure 12.16. The unequal bit error sensitivity of the codec becomes explicit in Figure 12.28, showing the segmental SNR of the G728 codec for channel BERs between 0.001% and 1%. The ten bits were arranged into these two groups based on the results shown in Figure 12.27 – bits 2,3,8,9 and 10 formed Class One and the other five bits formed Class Two. It can be seen that the Class One bits are about two or three times more sensitive than the Class Two bits, and therefore should be more strongly protected by the error correction and modulation schemes. For robustness reasons we have refrained from using a LTP.

We also investigated the error sensitivity of the 8 kbits/s mode of our low delay codec. LTP was not invoked, but the codec with a vector size of ten was used because, as was seen earlier, it gave a segmental SNR almost 2 dB higher than the 8 kbits/s mode of the codec with a constant vector size of five. As discussed in Section 12.7, the vector codebook entries for our codecs were trained as described in [345].

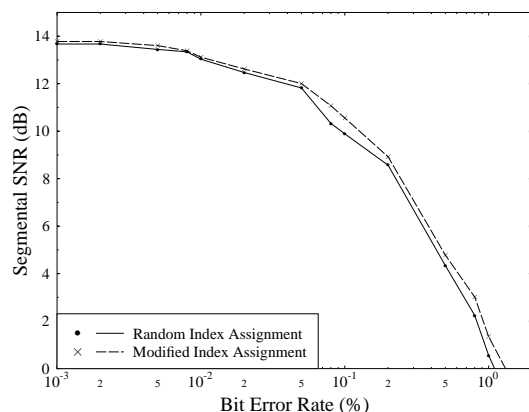


Figure 12.33: Segmental SNR of 8 kbits/s Codec Against Channel BER for Original and Rearranged Codebooks

However the 7 bit indices used to represent the 128 codebook entries are effectively randomly assigned. This assignment of indices to codebook entries does not affect the performance of the codec in error free conditions, but it is known that the robustness of vector quantizers to transmission errors can be improved by the careful allocation of indices to codebook entries [346]. This can be seen from Figure 12.33 which shows the segmental SNR of the 8 kbits/s codec for BERs between 0.001% and 1%. The solid line shows the performance of the codec using the codebook with the original index assignment, whereas the dashed line shows the performance of the codec when the index assignment was modified to improve the robustness of the codebook. A simple, non-optimum, algorithm was used to perform the index assignment and it is probable that the codec's robustness could be further improved by using a more effective minimisation algorithm such as simulated annealing. Also, as in the G728 codec, a natural binary code was used to represent the 8 quantized levels of the excitation gain. It is likely that the use for example of a Gray code to represent the 8 gain levels could also improve the codec's robustness.

The sensitivity of the ten bits used to represent each ten speech sample vector in our 8 kbits/s codec is shown in Figure 12.34. Again bits 1,2 and 3 are used to represent the excitation gain, and the other 7 bits represent the index of the codebook entry chosen to code the excitation shape. As in the case of the G728 codec the unequal error resilience of different bits can be clearly seen. Note in particular how the least significant of the 3 bits representing the excitation gain is much less sensitive than the 7 bits representing the codebook index, but that the two most sensitive gain bits are more sensitive than the codebook index bits.

Figure 12.35 shows the segmental SNR of the 8 kbits/s codec for BERs between 0.001% and 1%. Again the solid line shows the performance of the codec when the

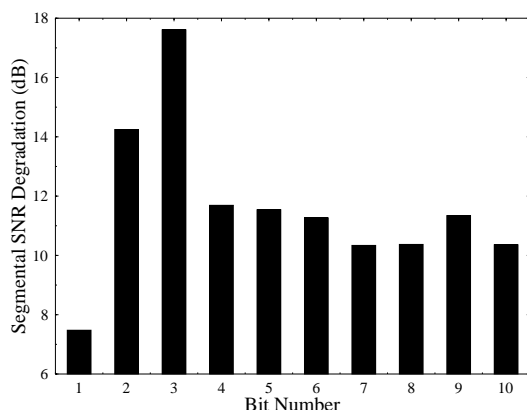


Figure 12.34: Degradation in 8 kbits/s Segmental SNR Caused by 10 % BER in Given Bits

errors are equally distributed amongst all ten bits, whereas the dashed lines show the performance when the errors are confined only to the 5 most sensitive Class One bits or the five least sensitive Class Two bits. The need for the more sensitive bits to be more protected by the FEC and modulation schemes is again apparent. These schemes, and how they are used to provide the required unequal error protection, is discussed in the next Section.

12.11.3 Transmission Issues

12.11.3.1 Higher-quality Mode

Based on the bit-sensitivity analysis presented in the previous Section we designed a sensitivity-matched transceiver scheme for both the higher and lower quality speech coding modes. Our basic design criterion was to generate an identical signalling rate in both modes in order to facilitate the transmission of speech within the same bandwidth, while providing higher robustness at a concomitant lower speech quality, if the channel conditions degrade.

Specifically, in the more vulnerable, higher-quality mode the 16-level Pilot Symbol Assisted Quadrature Amplitude Modulation (16-PSAQAM) scheme of Chapter 2 was used for the transmission of speech encoded at 16 kbps. In the more robust, lower-quality mode the 8 kbps encoded speech is transmitted using 4-PSAQAM at the same signalling rate. In our former work [52] we have found that typically it is sufficient to use a twin-class un-equal protection scheme, rather than more complex multi-class arrangements. We have also shown [47] that the maximum minimum distance square 16QAM constellation exhibits two different-integrity subchannels, namely the better quality C1 and lower quality C2 subchannels, where the bit error rate (BER) difference

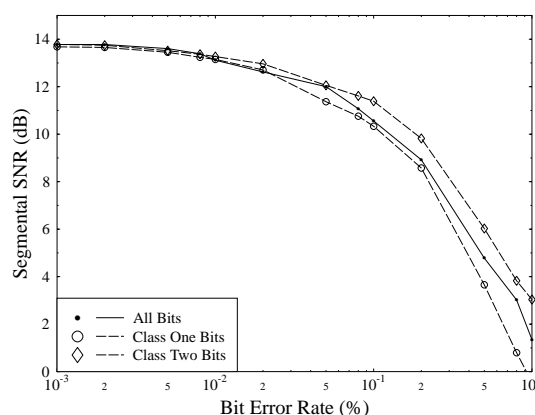


Figure 12.35: Segmental SNR of 8 kbits/s Codec Against Channel BER

is about a factor two in our operating Signal-to-Noise Ratio (SNR) range. This was also argued in Chapter 2.

Hence we would require a forward error correction (FEC) code of twice the correction capability for achieving a similar overall performance of both subchannels over Gaussian channels, where the errors have a typically random, rather than bursty distribution. Over bursty Rayleigh channels an even stronger FEC code would be required in order to balance the differences between the two subchannels. After some experimentation we opted for the binary Bose-Chaudhuri-Hocquenghem BCH(127,92,5) and BCH(124,68,9) codes of Chapter 4 for the protection of the 16 kbps encoded speech bits. The weaker code was used in the lower BER C1 subchannel and the stronger in the higher BER C2 16QAM subchannel. Upon evaluating the BERs of the coded subchannels over Rayleigh channels, which are not presented here due to lack of space, we found that a ratio of two in terms of coded BER was maintained.

Since the 16 kbps speech codec generated 160 bits/10ms frame, the 92 most vulnerable speech bits were directed to the better BCH(127,92,5) C1 16QAM subchannel, while the remaining 68 bits to the other subchannel. Since the C1 and C2 subchannels have an identical capacity, after adding some padding bits 128 bits of each subchannel were converted to 32 4-bit symbols. A control header of 30 bits was BCH(63,30,6) encoded, which was transmitted employing the more robust 4QAM mode of operation using 32 2-bit symbols. Finally, two ramp symbols were concatenated at both ends of the transmitted frame, which also incorporated four uniformly-spaced pilot symbols. A total of 104 symbols/10ms represented therefore 10 ms speech, yielding a signalling rate of 10.4 kBd. When using a bandwidth of 1728 kHz, as in the Digital European Cordless Telephone (DECT) system and an excess bandwidth of 50%, the multi-user signalling rate becomes 1152 kBd. Hence a total of $\text{INT}[1152/104]=110$ time-slots can be created, which allows us to support 55 duplex conversations in Time Division

Duplex (TDD) mode. The timeslot duration becomes $10\text{ms}/(110 \text{ slots}) \approx 90.091\mu\text{s}$.

12.11.3.2 Lower-quality Mode

In the lower-quality 8 kbps mode of operation 80bits/10ms are generated by the speech codecs, but the 4QAM scheme does not have two different integrity subchannels. Here we opted for the BCH(63,36,5) and BCH(62,44,3) codes in order to provide the required integrity subchannels for the speech codec. Again, after some padding the 64-bit coded subchannels are transmitted using 2-bit/symbol 4QAM, yielding 64 symbols. After incorporating the same 32-symbol header block, 4 ramp and 4 pilot symbols, as in case of the higher-quality mode, we arrive at a transmission burst of 104 symbols/10ms, yielding an identical signalling rate of 10.4 kBd.

12.11.4 Speech Transceiver Performance

The SEGSNR versus channel SNR performance of the proposed multimode transceiver is portrayed in Figure 12.36 for both 10.4 kBd modes of operation. Our channel conditions were based on the DECT-like propagation frequency of 1.9 GHz, signalling rate of 1152 kBd and pedestrian speed of $1\text{m/s}=3.6 \text{ km/h}$, which yielded a normalised Doppler frequency of $6.3\text{Hz}/1152\text{kBd} \approx 5.5 \cdot 10^{-3}$. Observe in the Figure that unimpaired speech quality was experienced for channel SNRs in excess of about 26 and 18 dBs in the less and more robust modes, respectively. When the channel SNR degrades substantially below 22 dB, it is more advantageous to switch to the inherently lower quality, but more robust and essentially error-free speech mode, demonstrating the advantages of the multimode concept. The effective single-user simplex bandwidth is $1728\text{kHz}/110 \text{ slots} \approx 15.71 \text{ kHz}$, while maintaining a total transmitter delay of 10 ms. Our current research is targeted at increasing the number of users supported using Packet Reservation Multiple Access.

12.12 Chapter Conclusions

In this chapter we highlighted the operation of the CCITT G728 16 kbps standard codec and proposed a range of low delay coding schemes operating between 16-8 and 8-4 kbits/s. While in the higher bitrate range entirely backward-adaptive predictive arrangements were used, in the lower range codecs using both forward and backward adaption of the long term filter have been considered, but all the codecs use backward adaption of the short term synthesis filter and so have frame sizes of at most 5 ms. Both relatively small trained shape codebooks and large algebraic codebooks were used. We found that the resulting codecs offered a range of reconstructed speech qualities between communications quality at 4 kbits/s to near-toll quality at 8 kbits/s. Lastly, an application example was given, demonstrating the practical applicability of the codecs portrayed. Let us now concentrate our attention on high-quality wideband speech compression in the next chapter.

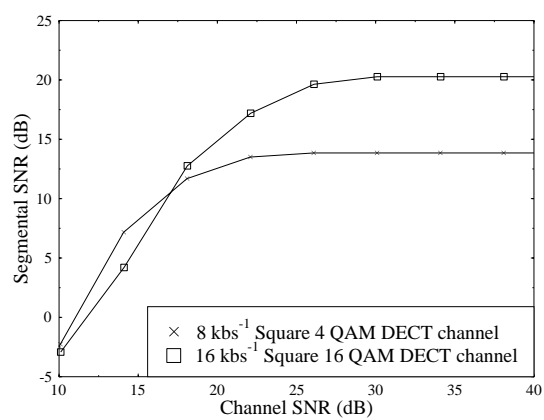
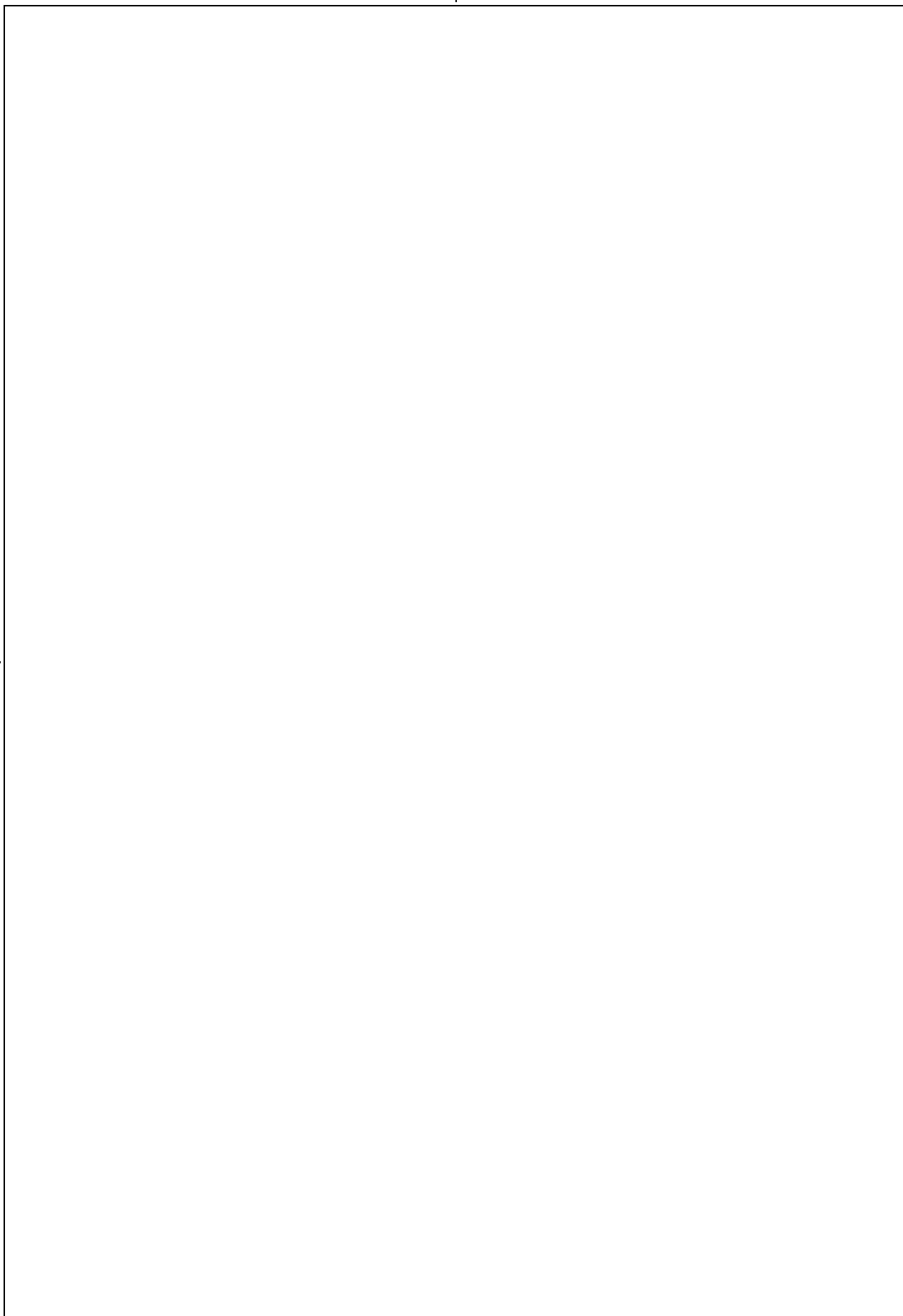


Figure 12.36: Segmental SNR versus Channel SNR Performance of the Proposed Multi-mode Transceiver

Part IV

Wideband and Sub-4kbps Coding and Transmission



608

670

Chapter 17

Zinc Function Excitation

17.1 Introduction

This chapter introduces a prototype waveform interpolation (PWI) speech coder that uses zinc function excitation (ZFE) [411]. A PWI scheme operates by encoding one pitch period-sized segment, a prototype segment, of speech for each frame. The slowly evolving nature of speech permits PWI to reduce the transmitted bit rates, while smooth waveform interpolation at the decoder between the prototype segments maintains good synthesized speech quality. Figure 17.1(a) shows two 20ms frames of voiced speech, with a pitch period in both frames highlighted in each to demonstrate the slow waveform evolution of speech. The same pitch periods are again highlighted for the LPC STP residual waveform, in Figure 17.1(b), demonstrating that PWI can also be used on the residual signal. Finally, Figure 17.1(c) displays the frequency spectrum for both frames, showing the evolution of the speech waveform in the frequency domain. The excitation waveforms employed in this chapter are the zinc basis functions [411], which efficiently model the LPC STP residual while reducing the speech's 'buzziness' when compared with the classical vocoders of Chapter 15 [411]. The previously introduced schematic in Figure 14.13 portrays the encoder structure for the Interpolated Zinc Function Prototype Excitation (IZFPE), which has the form of a closed loop LPC based coding method with optimized ZFE prototype segments for the speech. A similar structure is used in the PWI-ZFE coder described in this chapter.

This chapter follows the basic outline of the IZFPE coder introduced by Hiotakakos and Xydeas [410], but some sections of the scheme have been developed further. The chapter begins with an overview of the PWI-ZFE scheme, detailing the operational scenarios of the arrangement. This is followed by the introduction of the zinc basis functions together with the optimization process at the encoder, where the wavelets of Chapter 16 are harnessed to reduce the complexity of the process. For voiced speech frames the pitch detector employed and the prototype segment selection process are described, with a detailed discussion of the interpolation process, where the param-

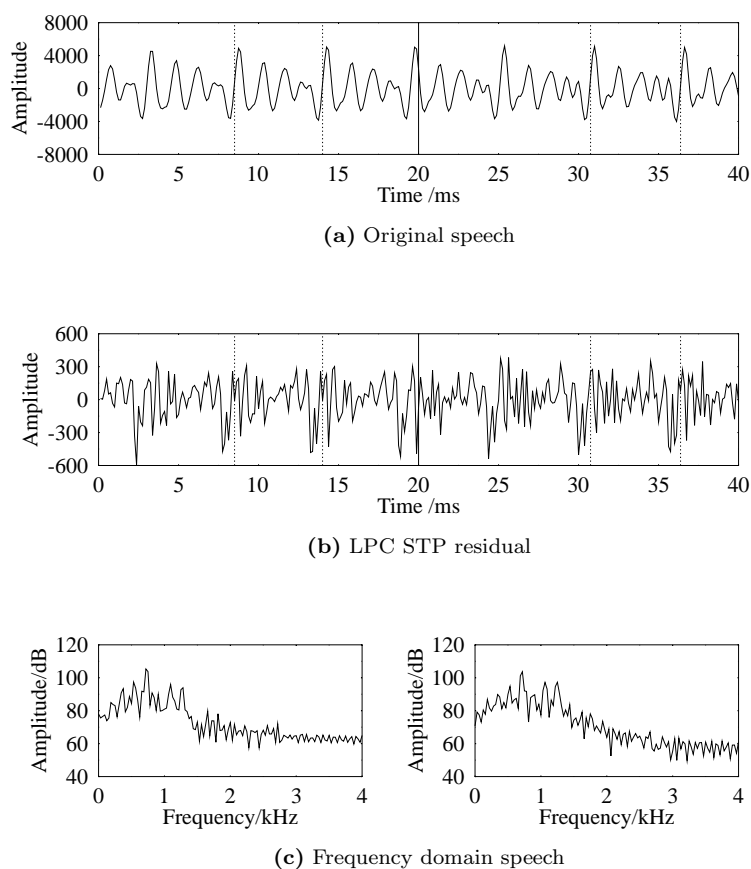


Figure 17.1: Two speech frames demonstrating the smoothly evolving nature of the speech waveform and that of the LPC STP residual in the time and frequency domain. The speech frames are from AF1, uttering the back vowel /ɔ/ in 'dog'

eters required for transmission are also given. Additionally, the unvoiced excitation and adaptive postfilter are briefly described. Finally, the performance of both a single ZFE and multiple ZFE arrangements are detailed.

17.2 Overview of Prototype Waveform Interpolation Zinc Function Excitation

This section gives an in depth description of the PWI-ZFE scheme, considering all possible operational scenarios at both the encoder and decoder. The number of coding scenarios is increased by the separate treatment of voiced and unvoiced frames, and also by the need to accurately represent the voiced excitation.

17.2.1 Coding Scenarios

For the PWI-ZFE encoder the current, the next and the previous two 20ms speech frames are evaluated, as shown in Figure 17.2, which is now described in depth. The knowledge of the four 20ms frames, namely frames $N + 1$, N , $N - 1$ and $N - 2$, is required in order to adequately treat voiced-unvoiced boundaries. It is these transition regions which are usually the most poorly represented speech segments in classical vocoders. The parameters encoded and transmitted during voiced and unvoiced periods are summarized towards the end of the chapter in Table 17.10, while the various coding scenarios are summarized in Table 17.1 and 17.2.

LPC STP analysis is performed for all speech frames and the RMS value is determined from the residual waveform. The pitch period of the speech frame is also determined. However, if the speech frame lacks any periodicity then the pitch period is assigned as zero and the speech frame is labelled as unvoiced. The various possible combinations of consecutive voiced (V) and unvoiced (U) frames are now considered.

17.2.1.1 U-U-U Encoder Scenario

If all the speech frames $N + 1$, N and $N - 1$ are classified as unvoiced, U-U-U, then the unvoiced parameters for frame N are sent to the decoder. The unvoiced parameters are the LPC coefficients, sent as LSFs, a voicing flag which is set to *off*, and the quantized RMS value of the LPC STP residual, as described in Table 17.1.

17.2.1.2 U-U-V Encoder Scenario

With a voicing sequence of U-U-V, where frame $N - 1$ is voiced, together with the unvoiced parameters an extra parameter b_s , the boundary shift parameter, must be conveyed to the decoder to be used for the voicing transition regions. In order to determine the voiced to unvoiced transition point, b_s , frame N is examined, searching for evidence of voicing, in segments sized by the pitch period. The boundary shift parameter b_s represents the number of pitch periods in the frame that contain voicing. At the decoder this voiced section of the predominantly unvoiced frame N is represented by the ZFE excitation reserved for the voiced segments.

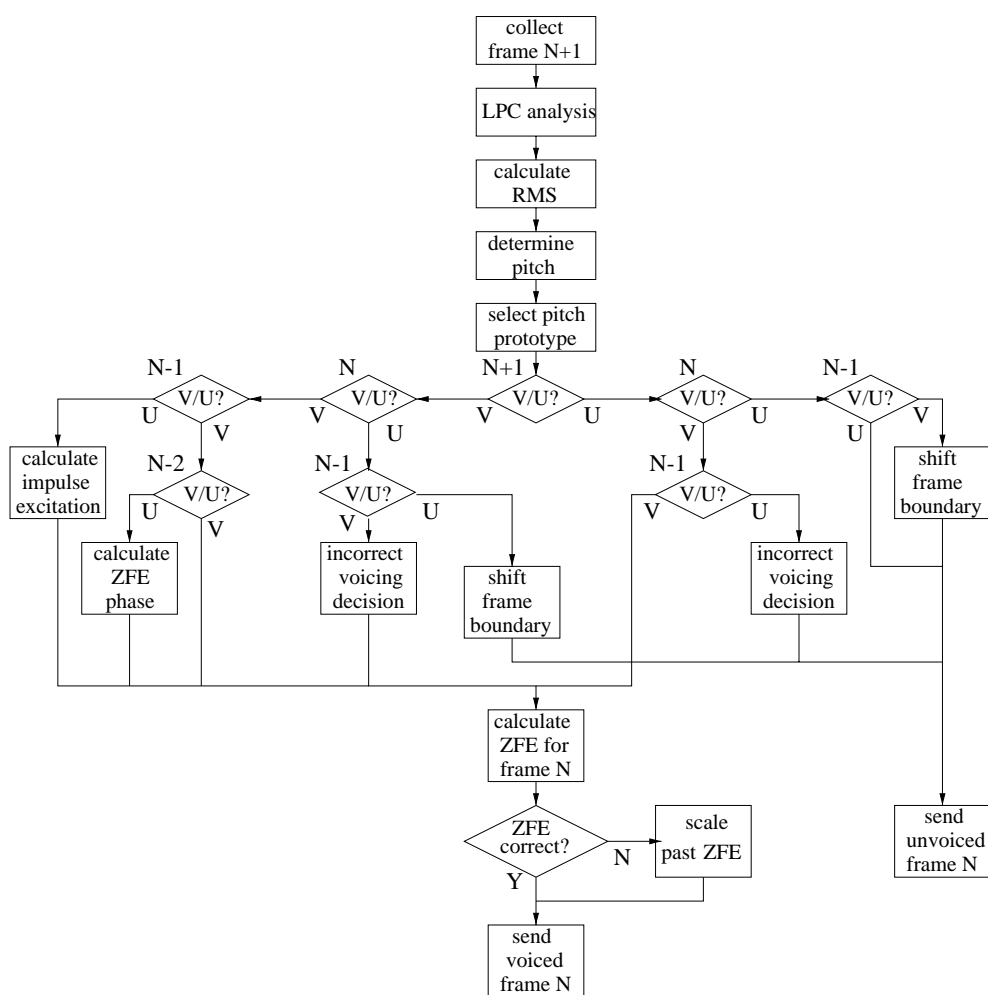


Figure 17.2: The encoder control structure for the PWI-ZFE arrangement.

**17.2. OVERVIEW OF PROTOTYPE WAVEFORM INTERPOLATION ZINC
FUNCTION EXCITATION****697**

N+1	N	N-1	Summary
U	U	U	Frame N is located in an unvoiced sequence. Quantize and transmit the RMS value of the LPC STP residual to the decoder.
U	U	V	A voiced-to-unvoiced transition boundary has been encountered. Calculate the section of frame N that is voiced and include this boundary shift parameter, b_s , in the transmission of frame N to the decoder.
V	U	U	An unvoiced-to-voiced transition boundary has been encountered. Calculate the section of frame N that is voiced and include this boundary shift parameter, b_s , in the transmission of frame N to the decoder.
U	V	U	Assume frame N should have been classified as unvoiced, hence, treat this scenario as an U-U-U sequence.
V	V	V	Frame N is situated in a voiced sequence. Calculate the ZFE parameters, A_1 , B_1 and λ_1 . Quantize the amplitude parameters A_1 and B_1 , and transmit parameters to decoder.
V	U	V	Assume frame N should have been labelled as voiced, hence, treat this case as a V-V-V sequence.
U	V	V	Treat this situation as a V-V-V sequence.
V	V	U	The start of a sequence of voiced frames has been encountered. Represent the excitation in the prototype segment with an impulse.

Table 17.1: Summary of encoder scenarios- see text for more detail.**17.2.1.3 V-U-U Encoder Scenario**

The boundary shift parameter is also sent for the voicing sequence V-U-U. However, for this sequence the predominantly unvoiced frame N is examined, in order to identify how many pitch period durations can be classified as voiced. The parameter b_s in frame N then represents the number of pitch periods in the unvoiced frame N that contain voicing. At the decoder this section of frame N is synthesized using voiced excitation.

17.2.1.4 U-V-U Encoder Scenario

A voicing sequence U-V-U is assumed to have an incorrect voicing decision in frame N . Hence, the voicing flag in frame N is set to zero and the procedure for an U-U-U sequence is followed.

17.2.1.5 V-V-V Encoder Scenario

For a voiced sequence of frames V-V-V the ZFE parameters for frame N are calculated. The ZFE is described by the position parameter λ_1 and the amplitude parameters A_1 and B_1 , as shown earlier in Figure 14.14. Further ZFE waveforms are shown in Figure 17.5, which also illustrates the definition of the ZFE phase referred to below. If frame $N - 2$ was also voiced, then the chosen ZFE is restricted by certain phase constraints, which will be detailed in Section 17.3, otherwise frame N is used to determine the phase restrictions. The selected ZFE represents a pitch-duration segment of the speech frame, which is referred to as the pitch prototype segment. If a ZFE that complies with the required phase restrictions is not found, then the ZFE parameters from frame $N - 1$ are scaled, in terms of the RMS energy of the respective frames, and then they are used in frame N . This is performed since it is assumed that the previous frame parameters will be an adequate substitute for frame N , due to the speech parameters slow time domain evolution. The parameters sent to the decoder include the LSFs and a voicing flag set to *on*. The ZFE parameters A_1 , B_1 and λ_1 , are required by the decoder to synthesize voiced speech, and the pitch prototype segment is defined by its starting point and the pitch period duration of the speech segment.

17.2.1.6 V-U-V Encoder Scenario

A voiced sequence of frames is also assumed for the voicing decisions V-U-V, with the frame N being assigned a pitch period half way between the pitch period for frame $N - 1$ and frame $N + 1$.

17.2.1.7 U-V-V Encoder Scenario

The voicing sequence U-V-V also follows the procedure of a V-V-V string, since the unvoiced decision of frame $N + 1$ is not considered until the V-U-V, or U-U-V scenarios.

17.2.1.8 V-V-U Encoder Scenario

The voicing decision V-V-U indicates that frame N will be the start of a voicing sequence. Frame $N + 1$, the second frame in the voicing sequence, typically constitutes a better reflection of the dynamics of the voiced sequence than the first one [410], hence the phase restrictions are determined from this frame. The first voiced frame, namely N , is represented by an excitation pulse similar to that used by the LPC vocoder of Chapter 15 [227].

The speech encoder introduces a delay of 40ms into the system, where the delay is caused by the necessity for frame $N + 1$ to verify voicing decisions. In the decoder control structure, shown in Figure 17.3, only the frames $N + 1$ and N are considered when synthesizing frame N , thus an additional 20ms delay is introduced.

17.2.1.9 U-V Decoder Scenario

If the sequence U-V occurs for the frame $N + 1$ and N respectively, then a voiced-to-unvoiced transition is encountered. Here the boundary shift parameter b_s , transmitted in frame $N + 1$, is multiplied by the pitch period in frame N , indicating the portion of

17.2. OVERVIEW OF PROTOTYPE WAVEFORM INTERPOLATION ZINC FUNCTION EXCITATION

699

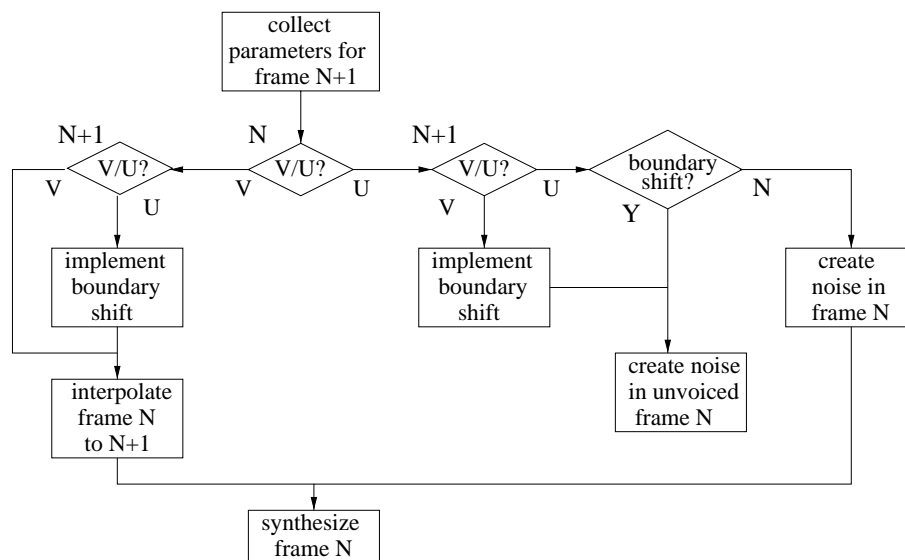


Figure 17.3: The decoder control structure for PWI-ZFE arrangement.

N+1	N	Summary
U	V	A voiced-to-unvoiced transition has been encountered. Label the portion of frame $N + 1$ that is voiced, and subsequently interpolate from the pitch prototype segment in frame N to the voiced sections in frame $N + 1$.
U	U	Frame N is calculated using a Gaussian noise excitation scaled by the RMS value for frame N .
V	U	An unvoiced-to-voiced transition has been encountered. Label the portion of frame N that is voiced and represent the relevant section of frame N by voiced excitation.
V	V	Interpolation is performed between the pitch prototype segments of frame N and frame $N + 1$.

Table 17.2: Summary of decoder scenarios- see text for more detail.

frame $N+1$ which was deemed voiced. The ZFE excitation for frame N is interpolated to the end of the voiced portion of frame $N+1$. Subsequently, the interpolation frame N is synthesized.

17.2.1.10 U-U Decoder Scenario

When the sequence U-U occurs for frame indices $N+1$ and N , if frame $N-1$ is unvoiced then frame N will be represented by a Gaussian noise excitation. However, if frame $N-1$ was voiced some of frame N will already be represented by a ZFE pulse. This will be indicated by the value of the boundary shift parameter b_s , thus only the unvoiced section of frame N is represented by Gaussian noise.

17.2.1.11 V-U Decoder Scenario

The sequence V-U indicates an unvoiced-to-voiced transition, hence, the value of the boundary shift parameter b_s conveyed by frame N is observed. Only the unvoiced section of frame N is represented by Gaussian noise, with the voiced portion represented by a ZFE interpolated from frame $N+1$.

17.2.1.12 V-V Decoder Scenario

The sequence V-V directs the decoder to interpolate the ZFE parameters between frame N and frame $N+1$. This interpolation process is described in Section 17.6, where it occurs for the region between pitch prototype segments. Thus each speech frame has its first half interpolated, while classified as frame $N+1$, with its second half interpolated during the next iteration, while classified as frame N .

Following this in depth description of the control structure of a PWI-ZFE scheme, as given by Figures 17.2 and 17.3, a deeper insight into the description of the ZFE is now given.

17.3 Zinc Function Modelling

The continuous zinc function used in the PWI-ZFE scheme to represent the LPC STP residual is defined by [411]:

$$z_k(t) = A_k \cdot \text{sinc}(t - \lambda_k) + B_k \cdot \text{cosc}(t - \lambda_k) \quad (17.1)$$

where $\text{sinc}(t) = \frac{\sin(2\pi f_c t)}{2\pi f_c t}$, $\text{cosc}(t) = \frac{1 - \cos(2\pi f_c t)}{2\pi f_c t}$, k is the k^{th} zinc function, A_k , B_k determine the amplitude of the zinc function and λ_k determines its location. For the discrete time case with a speech bandwidth of $f_c = 4\text{kHz}$ and a sampling frequency of $f_s = 8\text{kHz}$ we have [410]:

$$z_k(n) = A_k \cdot \text{sinc}(n - \lambda_k) + B_k \cdot \text{cosc}(n - \lambda_k) = \begin{cases} A_k & , n - \lambda_k = 0 \\ \frac{2B_k}{n\pi} & , n - \lambda_k = \text{odd} \\ 0 & , n - \lambda_k = \text{even} \end{cases} \quad (17.2)$$

17.3.1 Error Minimization

From Figure 14.16, which describes the analysis-by-synthesis process, the weighted error signal $e_w(n)$ can be described by:

$$e_w(n) = s_w(n) - \bar{s}_w(n) \quad (17.3)$$

$$= s_w(n) - m(n) - \left(\sum_{k=1}^K z_k(n) * h(n) \right) \quad (17.4)$$

$$= y(n) - \left(\sum_{k=1}^K z_k(n) * h(n) \right) \quad (17.5)$$

where $y(n) = s_w(n) - m(n)$, $m(n)$ is the memory of the LPC synthesis filter due to previous excitation segments, while $h(n)$ is the impulse response of the weighted synthesis filter, $W(z)$, and K is the number of ZFE pulses employed. Thus the error, $e_w(n)$, is the difference between the weighted original and weighted synthesized speech, with the synthesized speech being the ZFE passed through the synthesis filter, $W(z)$. This formulation of the error signal, where the filter's contribution is divided into filter memory $m(n)$ and impulse response $h(n)$, reduces the computational complexity required in the error minimization procedure. It is the Infinite Impulse Response (IIR) nature of the filter, which requires the memory to be considered in the error equation. For further details of the mathematics Chapter 3 of Steele [180] is recommended. The sum of the squared weighted error signal is given by:

$$E_w^{k+1} = \sum_{n=1}^{excint} (e_w^{k+1}(n))^2 \quad (17.6)$$

where $e_w^{k+1}(n)$ is the k^{th} order weighted error, achieved after k zinc basis functions have been modelled, and $excint$ is the length over which the error signal has to be minimized, here the pitch prototype segment length.

Appendix B describes the process of minimizing the squared error signal using Figure 14.16 and Equations 17.1 to 17.6. It is shown that the mean squared error signal is minimized if the expression:

$$\zeta_{mse} = \frac{R_{es}^2}{R_{ss}} + \frac{R_{ec}^2}{R_{cc}} \quad (17.7)$$

is maximized as a function of the ZFE position parameter λ_{k+1} , and:

$$R_{es} = \sum_{n=1}^{excint} (sinc(n - \lambda_{k+1}) * h(n)) \times e_w^k(n) \quad (17.8)$$

$$R_{ec} = \sum_{n=1}^{excint} (cosc(n - \lambda_{k+1}) * h(n)) \times e_w^k(n) \quad (17.9)$$

$$R_{ss} = \sum_{n=1}^{excint} (sinc(n - \lambda_{k+1}) * h(n))^2 \quad (17.10)$$

$$R_{cc} = \sum_{n=1}^{excint} (cosc(n - \lambda_{k+1}) * h(n))^2 \quad (17.11)$$

where $*$ indicates convolution.

Due to bit rate limitations it is now assumed that a single ZFE is used, *i.e.* $k = 1$, furthermore that the value *excint* becomes equivalent to the pitch period duration, with λ_k controlling the placement of the ZFE in the range [1 to *excint*].

The ZFE amplitude coefficients are given by Equations B.14 and B.15 of Appendix B, repeated here for convenience:

$$A_k = \frac{R_{es}}{R_{ss}} \quad (17.12)$$

$$B_k = \frac{R_{ec}}{R_{cc}} \quad (17.13)$$

The optimization involves computing ζ_{mse} in Equation 17.7 for all legitimate values of λ_1 in the range [1 to *excint*], subsequently finding the corresponding values for A_1 and B_1 from Equations 17.12 and 17.13. The computational complexity for this optimization procedure is now assessed.

17.3.2 Computational Complexity

The associated complexity is evaluated as follows and tabulated in Table 17.3. The calculation of the minimization criterion ζ_{mse} requires the highest computational complexity, where the convolution of both the *sinc* and *cosc* functions with the impulse response $h(n)$ is performed. From Equation 17.2 it can be seen that the *sinc* function is only evaluated when $n - \lambda_k = 0$, while the *cosc* function must be evaluated whenever $n - \lambda_k$ is odd. The convolved signals, involving the *sinc* and *cosc* signals, are then multiplied by the weighted error signal $e_w(n)$ to calculate R_{es} and R_{ec} , in Equations 17.8 and 17.9, respectively. Observing Equations 17.6 to 17.13, the computational complexity's dependence on the *excint* parameter can be seen. Thus, in Table 17.3 all values are calculated with the extreme values of *excint*, which are 20 and 147 the possible pitch period duration range in samples. The complexity increase is exponential, as shown in Figure 17.4 by the dashed line, where it can be seen that any pitch period longer than 90 samples in duration will exceed a complexity of 20 MFLOPS.

17.3.3 Reducing the Complexity of Zinc Function Excitation Optimization

The complexity of the ZFE minimization procedure can be reduced by considering the glottal closure instants (GCI) introduced in Chapter 16. In Chapter 16 wavelet analysis was harnessed to produce a pitch detector, where the pitch period was de-

17.3. ZINC FUNCTION MODELLING

703

Procedure	$excint = 20$ /MFLOPS	$excint = 147$ /MFLOPS
Convolve $sinc$ and $h(n)$	0.02	1.06
Convolve $cosc$ and $h(n)$	0.20	78.0
Calculate A_1	0.04	2.16
Calculate B_1	0.04	2.16
Total	0.3	83.38

Table 17.3: Computational complexity for error minimization in the PWI-ZFE encoder for the extremities of the $excint$ variable.

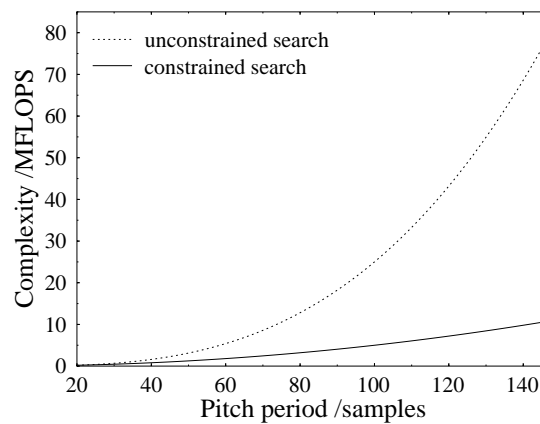


Figure 17.4: Computational complexity for the permitted pitch period range of 20 to 147 sample duration, for both an unrestricted and constrained search.

terminated as the distance between two GCIs. These GCIs indicate the snapping shut, or closure, of the vocal folds, which provides the impetus for the following pitch period. The energy peak caused by the GCI will typically be in close proximity to the position of the ZFE placed by the ZFE optimization process. This permits the possibility of reducing the complexity of the analysis-by-synthesis process. Figure 17.4 shows that as the number of possible ZFE positions increases linearly, the computational complexity increases exponentially. Hence, constraining the number of ZFE positions will ensure that the computational complexity remains at a realistic level. The constraining process is described next.

The first frame in a voiced sequence has no minimization procedure; simply, a single pulse is situated at the glottal pulse location within the prototype segment. For the other voiced frames, in order to maintain a moderate computational complexity, the number of possible ZFE positions is restricted as if the pitch period is always 20 samples. A suitable constraint is to have the ZFE located within 10 samples of the instant of glottal closure situated in the pitch prototype segment. Table 17.4 repeats the calculations of Table 17.3, for complexities related to 20 and 147 sample pitch periods, for a restricted search. In Figure 17.4 the solid line represents the computational complexity of a restricted search procedure in locating the ZFE. The maximum complexity for a 147 sample pitch period is 11MFLOPS. The degradation

Procedure	$excint = 20$ /MFLOPS	$excint = 147$ /MFLOPS
Convolve <i>sinc</i> and $h(n)$	0.02	0.15
Convolve <i>cosc</i> and $h(n)$	0.20	10.73
Calculate A_1	0.04	0.29
Calculate B_1	0.04	0.29
Total	0.30	11.46

Table 17.4: Computational complexity for error minimization in the PWI-ZFE encoder with a restricted search procedure.

to the speech coder's performance, caused by restricting the number of ZFE locations, is quantified in Section 17.4.2.

17.3.4 Phases of the Zinc Functions

There are four possible phases of the ZFE produced by four combinations of positive or negative valued A_1 and B_1 parameters, which is demonstrated in Figure 17.5 for parameter values of $A_1 = \pm 1$ and $B_1 = \pm 1$. Explicitly, if $|A_1| = 1$ and $|B_1| = 1$, then the possible phases of the ZFE are the following: $A_1 = 1$ $B_1 = 1$, $A_1 = 1$ $B_1 = -1$, $A_1 = -1$ $B_1 = 1$, and $A_1 = -1$ $B_1 = -1$. The phase of the ZFE is determined during the error minimization process, where the calculated A_1 , B_1 values of Equations 17.12 and 17.13 will determine the ZFE phase. It should be noted that for successful interpolation at the decoder the phase of the ZFE should remain constant throughout each voiced sequence.

Following this insight into zinc function modelling the practical formulation of an PWI-ZFE coder is discussed. Initially, the procedures requiring pitch period knowledge are discussed, which are followed by details of voiced and unvoiced excitation considerations.

17.4 Pitch Detection

The PWI-ZFE coder located the voiced frame's pitch period using the autocorrelation-based wavelet pitch detector described in Section 16.5.2, which has a computational complexity of 2.67 MFLOPS. This Section investigates methods of making voiced-unvoiced decisions for pitch-sized segments, and methods for identifying a pitch period segment.

17.4.1 Voiced-Unvoiced Boundaries

Classifying a segment of speech as voiced or unvoiced is particularly difficult at the transition regions, hence a segment of voiced speech can easily become classified as unvoiced. Thus, in the transition frame pitch-duration sized segments are examined for evidence of voicing. In this case the autocorrelation approach cannot be used, as several pitch periods are not available for the correlation procedure. Instead a side

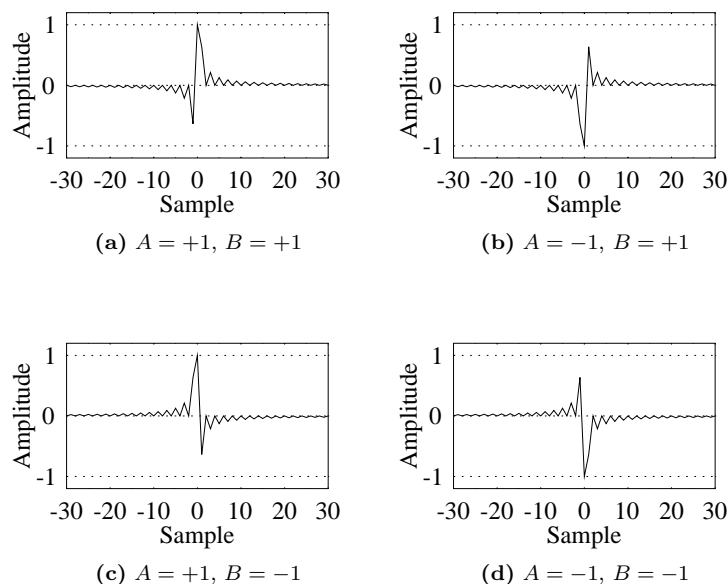


Figure 17.5: The four different phases possible for the ZFE waveform of Equation 17.1.

result of the wavelet based pitch detector is utilized, namely that for every speech frame candidate glottal pulse locations exist.

Therefore, if the first voiced frame in a voiced sequence is frame N , then frame $N - 1$ is examined for boundary shift. If a periodicity, close to the pitch period of frame N exists over an end portion of frame $N - 1$, this end portion of frame $N - 1$ is designated as voiced. Similarly, if the final voiced frame in a voiced sequence is frame N , then frame $N + 1$ is examined for boundary shift. Any starting portion of frame $N + 1$ that has periodicity close to the pitch period of frame N is declared voiced.

In the speech decoder it is important for the ZFE parameters to be interpolated over an integer number of pitch periods. Thus, the precise duration of voiced speech in the transition frame is not completely defined until the λ_1 interpolation process, to be described in Section 17.6, is concluded.

17.4.2 Pitch Prototype Selection

For each speech frame classed as voiced a prototype pitch segment is located, parameterized, encoded and transmitted. Subsequently, at the decoder interpolation between adjacent prototypes is performed. For smooth waveform interpolation the prototype must be a pitch period in duration, since this speech segment captures all elements of the pitch period cycle, thus enabling a good reconstruction of the original speech.

The prototype selection for the first voiced frame is demonstrated in Figure 17.6. If P is the pitch period of the voiced frame, then P samples in the centre of the

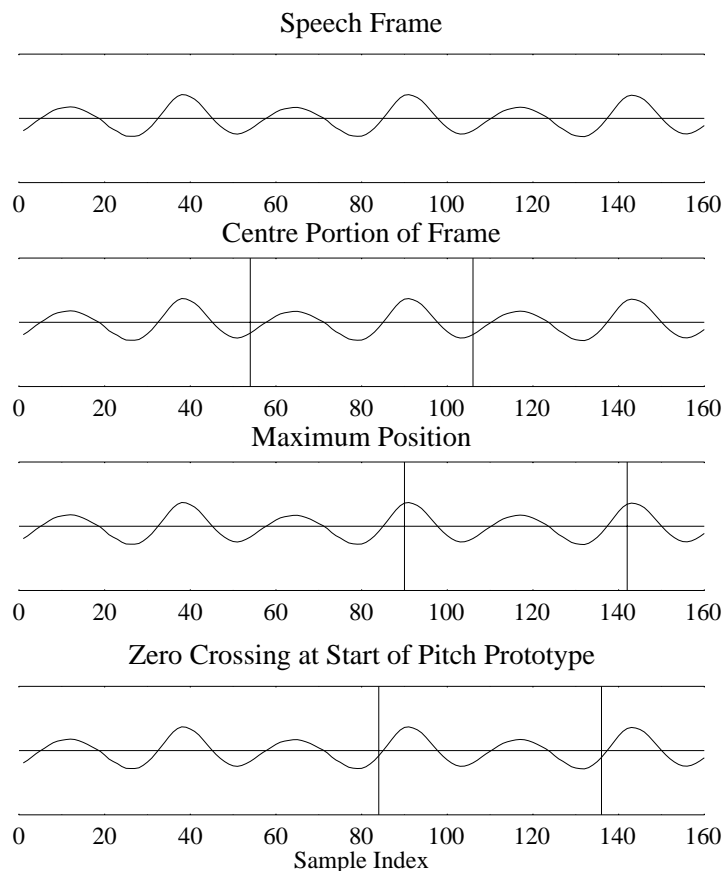


Figure 17.6: Pitch prototype selection for AM2 uttering the nasal consonant /n/ from ‘end’.

frame are selected as the initial prototype selection, as shown in the second trace of Figure 17.6. Following Hiotakakos and Xydeas [410], the maximum amplitude is found in the frame, as shown in the middle trace of Figure 17.6. Finally, the zero-crossing immediately to the left of this maximum is selected as the start of the pitch prototype segment, as indicated at the bottom of the Figure. The end of the pitch prototype segment is a pitch period duration away. Locating the start of the pitch prototype segment near a zero crossing helps to reduce discontinuities in the speech encoding process.

It is also beneficial in the interpolation procedure of the decoder if consecutive ZFE locations are smoothly evolving. Therefore, close similarity between consecutive prototype segments within a voiced sequence of frames is desirable. Thus, after the first frame the procedure of Hiotakakos and Xydeas [410] is no longer followed. Instead the cross-correlation between consecutive pitch prototype segments [402] of the other speech frames is performed. These subsequent pitch prototype segments are calculated from the maximum cross-correlation between the current speech frame and previous

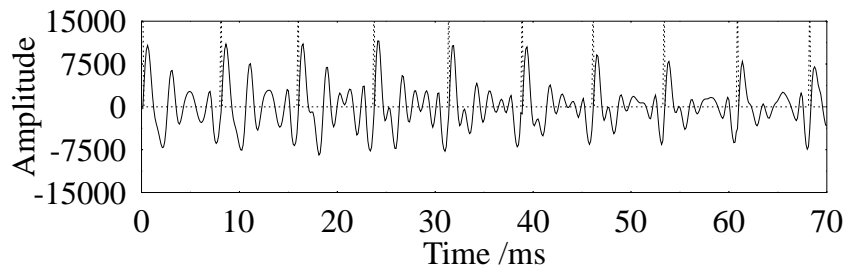


Figure 17.7: Concatenated speech signal prototype segments producing a smoothly evolving waveform. The dotted lines represent the prototype boundaries.

	unconstrained search	constrained search
no phase restrictions	3.36dB	2.68dB
phase restrictions	2.49dB	1.36dB

Table 17.5: SEGSNR results for the optimization process with and without phase restrictions, or a constrained search.

pitch prototype segment. Figure 17.7 shows how, at the encoder, the speech waveform prototype segments can be concatenated to produce a smoothly evolving waveform.

In order to further improve the probability that consecutive ZFEs have similar locations within their prototype segments, any instants of glottal closure that are not close to the previous segment's ZFE location are discarded, with the previous ZFE location used to search for the new ZFE in the current prototype segment.

At the encoder the introduction of constraining the location of the ZFE pulse, to within ± 10 positions, reduces the SEGSNR value, as shown in Table 17.5. The major drawback of the constrained search is the possibility that the optimization process is degraded through the limited range of ZFE locations searched. Additionally, it is possible to observe the degradation to the Mean Squared Error (MSE) optimization, caused by the phase restrictions imposed on the ZFEs and detailed in Section 17.3.4. Table 17.5 displays the SEGSNR values of the concatenated voiced prototype speech segments. The unvoiced segments are ignored, since these speech spurts are represented by noise, thus a SEGSNR value would be meaningless.

Observing Table 17.5 for a totally unconstrained search, the SEGSNR achieved by the ZFE optimization loop is 3.36dB. The process of either implementing the above-mentioned ZFE phase restriction or constraining the permitted ZFE locations to the vicinity of the GCIs reduces the voiced segments' SEGSNR after ZFE optimization by 0.87dB and 0.68dB, respectively. Restricting both the phase and the ZFE locations reduces the SEGSNR by 2dB. However, in perceptual terms the ZFE interpolation procedure, described in Section 17.6, actually improves the subjective quality of the decoded speech due to the smooth speech waveform evolution facilitated, despite the SEGSNR degradation of about 0.87dB caused by imposing phase restrictions. Similarly, the extra degradation of about 1.13dB caused by constraining the location

of the ZFEs also improves the perceived decoded speech quality due to smoother waveform interpolation.

17.5 Voiced Speech

For frames designated as voiced the excitation signal is a single ZFE. For a single ZFE the equations defined in Section 17.3 and Appendix B are simplified, since the k^{th} stage error Equation 17.5 becomes:

$$e_w^0(n) = y(n) \quad (17.14)$$

Therefore, Equation 17.6 for the weighted error of a single ZFE is given by:

$$E_w^1(n) = \left(\sum_{n=1}^{excint} e_w^1(n) \right)^2 \quad (17.15)$$

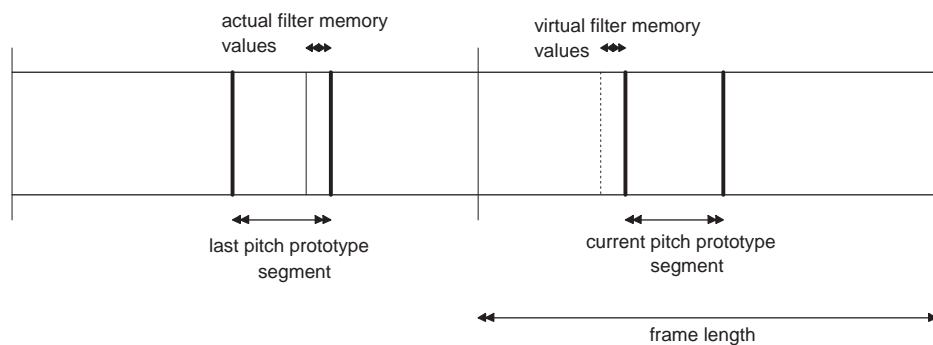
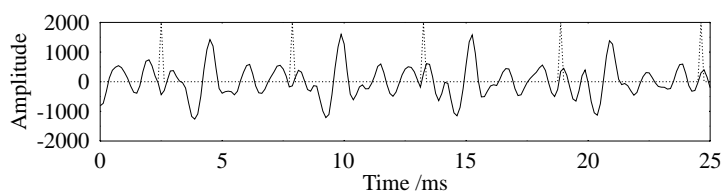
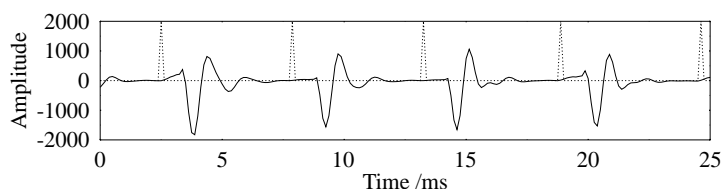
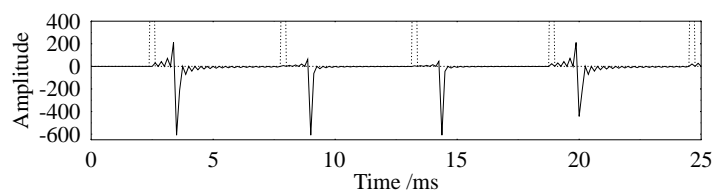
where $e_w^1(n) = y(n) - [z(n) * h(n)]$. Equations 17.8 and 17.9 are simplified to:

$$R_{es} = \sum_{n=1}^{excint} (\text{sinc}(n - \lambda_1) * h(n)) \times y(n) \quad (17.16)$$

$$R_{ec} = \sum_{n=1}^{excint} (\text{cosc}(n - \lambda_1) * h(n)) \times y(n) \quad (17.17)$$

Calculating the ZFE, which best represents the pitch prototype, involves locating the value of λ_1 between 0 and the pitch period that maximizes the expression for ζ_{mse} given in Equation 17.7. While calculating ζ_{mse} , $h(n)$ is the impulse response of the weighted synthesis filter $W(z)$, and the weighted error signal e_w is the LPC residual signal minus the LPC STP filter's memory, as shown by Equation 17.14. The use of prototype segments produces a ZFE determination process that is a discontinuous task, thus the actual filter memory is not explicitly available for the ZFE optimization process. Subsequently the filter's memory is assumed to be due to the previous ZFE. Figure 17.8 shows two consecutive speech frames, where the previous pitch prototype segment has its final p samples highlighted as LPC synthesis filter memory values, while for the current pitch prototype segment these p samples constitute virtual filter memory. Thus, for the error minimization procedure the speech between the prototype segments has been effectively removed.

Once the value of λ_1 , that produces the maximum ζ_{mse} value, has been determined, the appropriate values of A_1 and B_1 are calculated using Equations 17.12 and 17.13. Figure 17.7 displayed the smooth evolution of the concatenated pitch prototype segments. If the ZFEs selected for these prototype segments are passed through the weighted LPC STP synthesis filter, the resulting waveform should be a good match for the weighted speech waveform used in the minimization process. This is shown in Figure 17.9, characterizing the analysis-by-synthesis approach used in the PWI-ZFE encoder.

**Figure 17.8:** Determining the LPC filter memory.**(a)** Prototype segments of the weighted speech**(b)** Synthesized prototypes for the weighted speech**(c)** Zinc function excitation**Figure 17.9:** Demonstrating the process of analysis-by-synthesis encoding for prototype segments that have been concatenated to produce a smoothly evolving waveform. The dotted spikes indicate the boundaries between prototype segments.

The above procedure is only followed for the phase constraining frame, for subsequent frames in a voiced sequence the ZFE selected must have the phase dictated by the phase constraining frame. If phase restrictions are not followed, then during the interpolation process a change in the sign of A_1 or B_1 will result in some small valued interpolated ZFEs as the values pass through zero. For each legitimate zinc pulse position, λ_1 , the sign of A_1 and B_1 are initially checked, where the value of ζ_{mse} is calculated only if the phase restriction is satisfied. Therefore, the maximum value of ζ_{mse} associated with a suitably phased ZFE is selected as the excitation signal. It is feasible that a suitably phased ZFE will not be found, indeed with the test database 13% of the frames did not have a suitable ZFE. If this occurs, then the previous ZFE is scaled, as explained below, and used for the current speech frame. The scaling is based on the RMS value of the LPC residual after STP analysis which is defined by:

$$A_1(N) = \delta_s A_1(N-1) \quad (17.18)$$

$$B_1(N) = \delta_s B_1(N-1) \quad (17.19)$$

where

$$\delta_s = \frac{\text{RMS of LPC residual } N}{\text{RMS of LPC residual } N-1} \quad (17.20)$$

The value of $\lambda_1(N)$ is assigned to be the ZFE position in frame $N-1$, becoming $\lambda_1(N-1)$.

17.5.1 Energy Scaling

The values of A_1 and B_1 determined in the voiced speech encoding process produce an attenuation in the signal level from the original prototype signal. The cause of this attenuation is due to the nature of the minimization process described in Section 17.3, where the best waveform match between the synthesized and original speech is found. However, the minimization process does not consider the relative energies of the original weighted waveform and the synthesized weighted waveform. Thus, the values of the A_1 and B_1 parameters are scaled to ensure that the energies of the original and reconstructed prototype signals are equal, requiring that:

$$\sum_{n=1}^{excinct} (z(n) * h(n))^2 = \sum_{n=1}^{excinct} (\bar{s}_w(n) - m(n))^2 \quad (17.21)$$

where $h(n)$ is the impulse response of the weighted LPC STP synthesis filter, $\bar{s}_w(n)$ is the weighted speech signal, and $m(n)$ is the memory of the weighted LPC STP synthesis filter. Ideally, the energy of the excitation signals will also be equal, thus:

$$\sum_{n=1}^{excint} z(n)^2 = \sum_{n=1}^{excint} r(n)^2 \quad (17.22)$$

where $r(n)$ is the LPC STP residual.

The above equation shows that it is desirable to ensure that the energy of the synthesized excitation is equal to the energy of the LPC STP residual for the prototype

Quantizer Scheme	SNR /dB for A_1	SNR /dB for B_1
4-bit	10.45	10.67
5-bit	18.02	19.77
6-bit	26.47	27.07

Table 17.6: SNR values for SQ of the A_1 and B_1 parameters.

segment. Upon, expanding the left hand side of Equation 17.22 to include A_1 and B_1 , also introducing the scale factor S_{AB} that will ensure Equation 17.22 is obeyed, we have:

$$\sum_{n=1}^{excinct} [\sqrt{S_{AB}}A_1 \sin c(n - \lambda_1) + \sqrt{S_{AB}}B_1 \cos c(n - \lambda_1)]^2 = \sum_{n=1}^{excinct} r(n)^2 \quad (17.23)$$

where,

$$S_{AB} = \frac{\sum_{n=1}^{excinct} r(n)^2}{\sum_{n=1}^{excinct} [A_1 \sin c(n - \lambda_1) + B_1 \cos c(n - \lambda_1)]^2} \quad (17.24)$$

Here the factor S_{AB} represents the difference in energy between the original and synthesized excitation. Thus, by multiplying both the A_1 and B_1 parameters by $\sqrt{S_{AB}}$ the energies of the synthesized and original excitation prototype segments will match.

17.5.2 Quantization

Once the A_1 and B_1 parameters have been determined they must be quantized. The Max-Lloyd quantizer, described in Section 15.4, requires knowledge of the A_1 and B_1 parameters' PDF, which are shown in Figure 17.10, where the PDF is generated from the unquantized A_1 and B_1 parameters of the training speech database, described in Section 14.4.

The Max-Lloyd quantizer was used to create 4.5 and 6-bit SQs for both the A_1 and B_1 parameters. Table 17.6 shows the SNR values for the A_1 and B_1 parameters for the various quantization schemes.

In order for further insight into the performance of the various quantizers, the SEGSR and SD measures were calculated for the synthesized and original speech prototype segments. Together with the quantized A_1 and B_1 values the SEGSR and SD measures were calculated for the unquantized A_1 and B_1 values. Table 17.7 shows the SEGSR values achieved. While low, the SEGSR values demonstrate that the 6-bit quantization produces a SEGSR performance similar to the unquantized parameters.

Table 17.8 shows the SD values achieved, which demonstrate again that the 6-bit quantizers produce little degradation. The 6-bit A_1 and B_1 SQs were selected due to their transparency in the SEGSR and SD tests. They have SNR values of 26.47dB and 27.07dB, respectively, as seen in Table 17.6.

The interpolation of the voiced excitation performed at the decoder is described next, where pitch synchronous interpolation of the ZFE and LSFs are implemented.

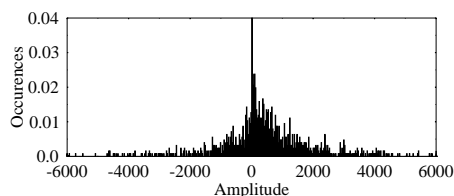
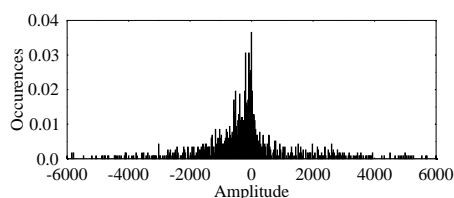
(a) PDF for the ZFE A parameter(b) PDF for the ZFE B parameter

Figure 17.10: Graphs for the PDFs of a) A_1 and b) B_1 ZFE parameters, created from the combination of A_1 and B_1 parameters from 45 seconds of speech.

Quantizer Scheme	SEGSNR /dB
unquantized	1.36
4-bit	0.21
5-bit	1.00
6-bit	1.29

Table 17.7: SEGSNR values between the original and synthesized prototype segments for a selection of SQs for the A_1 and B_1 parameters.

Quantizer Scheme	SD /dB
unquantized	4.53
4-bit	4.90
5-bit	4.60
6-bit	4.53

Table 17.8: SD values for the synthesized prototype segments for a selection of SQs for the A_1 and B_1 parameters.

17.6 Excitation Interpolation Between Prototype Segments

Having determined the prototype segments for the adjacent speech frames, interpolation is necessary in order to provide a continuous excitation signal between them. The interpolation process is investigated in this Section.

17.6.1 ZFE Interpolation Regions

The associated interpolation operations will be first stated in general terms, subsequently, using the equations derived and the parameter values of Table 17.9, they will be augmented using a numerical example. We also refer forward to traces three and four of Figures 17.11 and 17.12, which portray the associated interpolation operations.

Initially we follow the method of Hiotakakos and Xydeas [410] with interpolation performed over an interpolation region d_{pit} , where d_{pit} contains an integer number of pitch periods. The provisional interpolation region, d'_{pit} , which may not contain an integer number of pitch periods, begins at the start of the prototype segment in frame $N - 1$ and finishes at the end of the prototype segment in frame N . The number of pitch synchronous intervals, N_{pit} , between the two prototype regions is given by the ratio of the provisional interpolation region to the average pitch period during this region [410]:

$$N_{pit} = nint\left\{\frac{2d'_{pit}}{P(N) + P(N - 1)}\right\} \quad (17.25)$$

where $P(N)$ and $P(N - 1)$ represent the pitch period in frames N and $N - 1$ respectively, and $nint$ signifies the nearest integer. If $P(N)$ and $P(N - 1)$ are different, then the smooth interpolation of the pitch period over the interpolation region is required. This is achieved by calculating the average pitch period alteration necessary to convert $P(N - 1)$ to $P(N)$ over N_{pit} pitch synchronous intervals, where the associated pitch interpolation factor ϵ_{pit} is defined as [410]:

$$\epsilon_{pit} = \frac{P(N) - P(N - 1)}{N_{pit} - 1} \quad (17.26)$$

The final interpolation region, d_{pit} , is given by the sum of the pitch periods over the interpolation region constituted by N_{pit} number of pitch period intervals [410]:

$$d_{pit} = \sum_{np=1}^{N_{pit}} p(n_p) \quad (17.27)$$

where $p(n_p)$ are the pitch period values between the $P(N - 1)$ and $P(N)$, with $p(n_p) = P(N - 1) + (n_p - 1) \cdot \epsilon_{pit}$ and $n_p = 1..N_{pit}$. In general the start and finish of the prototype region in frame N will be altered by the interpolation process, since the provisional interpolation region d'_{pit} is generally extended or shortened, to become the interpolation region d_{pit} . To ensure correct operation between frame N and frame

$N - 1$ the change in the prototype position must be noted:

$$change = d'_{pit} - d_{pit} \quad (17.28)$$

and then we assign $start(N) = start(N) - change$, where $start(N)$ is the beginning of the prototype segment in frame N . Thus, the start of the prototype segment in frame N together with the position of the ZFE parameter λ_1 within the frame are altered, in order to compensate for the changes to the interpolation region. Maintaining the position parameter λ_1 at the same location of the prototype segment sustains the shape within the prototype excitation, but introduces a time misalignment with the original speech, where this time misalignment has no perceptual effect.

17.6.2 ZFE Amplitude Parameter Interpolation

The interpolated positions for the ZFE amplitude parameters are given by [410]:

$$A_{1,n_p} = A_1(N-1) + (n_p - 1) \frac{A_1(N) - A_1(N-1)}{N_{pit} - 1} \quad (17.29)$$

$$B_{1,n_p} = B_1(N-1) + (n_p - 1) \frac{B_1(N) - B_1(N-1)}{N_{pit} - 1} \quad (17.30)$$

where the formulae reflect a linear sampling of the A_1 and B_1 parameters between the adjacent prototype functions. Explicitly, given the starting value $A_1(N-1)$ and the difference $\Delta_{pit} = A_1(N) - A_1(N-1)$ the corresponding gradient is $[\Delta_{pit}/N_{pit} - 1]$, where N_{pit} is the number of pitch synchronous intervals between $A_1(N)$ and $A_1(N-1)$ allowing us to calculate the appropriate values A_{1,n_p} .

17.6.3 ZFE Position Parameter Interpolation

Interpolating the position of the ZFEs in a similar manner to their amplitudes does not produce a smoothly evolving excitation signal. Instead, the pulse position within each prototype segment is kept stationary throughout a voiced sequence. This introduces time misalignment between the original and synthesized waveforms, but maintains a smooth excitation signal. In order to compensate for changes in the length of prototype segments the normalized location of the initial ZFE position is calculated according to:

$$\lambda_r = \frac{\lambda_1(N)}{P(N)} \quad (17.31)$$

where $P(N)$ is the pitch period of the first frame in the voiced frame sequence. For all subsequent frames in the voiced sequence the position of the ZFE is calculated by:

$$\lambda_1(N) = nint\{\lambda_r * P(N)\} \quad (17.32)$$

where $nint\{\cdot\}$ represents rounding to the nearest integer.

For the sake of illustration the interpolation process is followed below for the two speech frames whose parameters are described in Table 17.9. The initial provisional interpolation region commences at the beginning of the prototype segment in frame

17.6. EXCITATION INTERPOLATION BETWEEN PROTOTYPE SEGMENTS

Speech Frame	Pitch Period	Zero-Crossing	A_1	B_1	λ_1
$N - 1$	52	64	-431	186	16
N	52	56	-573	673	20

Table 17.9: Transmitted parameters for voiced speech.

$N - 1$ and finishes at the end of the prototype segment in frame N . Since the zero crossing in frame $N - 1$ is at sample index 64 the provisional interpolation region in frame $N - 1$ is of duration $(160 - 64)$, while in frame N it finishes one pitch period duration, namely 52 samples, after the zero crossing at position 56, yielding:

$$d'_{pit} = (160 - 64) + (56 + 52) = 204$$

Using Equation 17.25 the number of pitch synchronous intervals, between the two consecutive prototype segments in frames N and $N - 1$, is given by d'_{pit} divided by the average pitch period duration of $[P(N) + P(N - 1)]/2$, yielding:

$$N_{pit} = \text{nint}\left\{\frac{2 \times 204}{52 + 52}\right\} = 4$$

As $P(N)$ and $P(N - 1)$ are identical the pitch interpolation factor ϵ_{pit} , of Equation 17.26, will be zero, while the interpolation region containing $N = 4$ consecutive pitch periods and defined by Equation 17.27 becomes:

$$d_{pit} = \sum_{n_p=1}^4 52 = 208$$

The interpolated ZFE magnitudes and positions can then be calculated using the parameters in Table 17.9 and Equations 17.29 to 17.32 for frame $N - 1$, the first voiced frame in the sequence, yielding:

$$\begin{aligned} A_{1,n_p} &= -431 + n_p \times \frac{-573 + 431}{3} = -478; -526; -573; \\ B_{1,n_p} &= 186 + n_p \times \frac{673 - 186}{3} = 348; 511; 673; \\ \lambda_r &= \frac{16}{52} = 0.308 \\ \lambda_1(N) &= 0.308 * 52 = 16 \end{aligned}$$

Again, the associated operations are illustrated in traces three and four of Figure 17.11 and 17.12.

17.6.4 Implicit Signalling of Prototype Zero Crossing

In order to perform the interpolation procedure described above the zero-crossing parameter of the prototype segments must be transmitted to the decoder. However, it can be observed that the zero-crossing values of the prototype segments are approximately a frame length apart, thus following the principle of interpolating between prototype segments in each frame. Hence, instead of explicitly transmitting the zero-

crossing parameter, it can be assumed that the start of the prototype segments are a frame length apart. An arbitrary starting point for the prototype segments could be $FL/2$, where FL is the speech frame length.

Using this scenario, the interpolation procedure example of Section 17.6.3 is repeated with both zero-crossings set to 80. The initial provisional interpolation region is calculated as:

$$d'_{pit} = (160 - 80) + (80 + 52) = 212 \quad (17.33)$$

The number of pitch synchronous intervals is given by:

$$N_{pit} = \text{rint}\left\{\frac{2 \times 212}{52 + 52}\right\} = 4 \quad (17.34)$$

Thus, the interpolation region defined by Equation 17.27 will become:

$$d_{pit} = \sum_{np=1}^4 52 = 208 \quad (17.35)$$

yielding the same distance as in the example of Section 17.6.3, where the zero-crossing value was explicitly transmitted. Hence, it is feasible not to transmit the zero-crossing location to the decoder. Indeed, the assumption of a zero-crossing value of 80 had no perceptual effect on speech quality at the decoder.

17.6.5 Removal of ZFE Pulse Position Signalling and Interpolation

In the λ_1 transmission procedure, although λ_1 is transmitted every frame only the first λ_1 in every voiced sequence is used in the interpolation process, thus, λ_1 is predictable and hence it contains much redundancy. Furthermore, when constructing the excitation waveform at the decoder every ZFE is permitted to extend over three interpolation regions, namely, its allotted region together with the previous and the next region. This allows ZFEs near the interpolation region boundaries to be fully represented in the excitation waveform, while ensuring that every ZFE will have a tapered low energy value when it is curtailed. It is suggested that the true position of the ZFE pulse, λ_1 , is arbitrary and need not be transmitted. Following this hypothesis, our experience shows that we can set $\lambda_1 = 0$ at the decoder, which has no audible degrading effect on the speech quality.

17.6.6 Pitch Synchronous Interpolation of Line Spectrum Frequencies

The LSF values can also be interpolated on a pitch synchronous basis, following the approach of Equations 17.29 and 17.30, giving:

$$LSF_{i,n} = LSF_i(N-1) + (n_p - 1) \frac{LSF_i(N) - LSF_i(N-1)}{N_{pit} - 1} \quad (17.36)$$

where $LSF_i(N-1)$ is the previous i^{th} LSF and $LSF_i(N)$ is the current i^{th} LSF.

17.6.7 ZFE Interpolation Example

An example of the ZFE excitation reconstructing the original speech is given in Figure 17.11, which is a speech waveform from the testfile AF2. Following the steps of the encoding and decoding process in the Figure, initially a pitch prototype segment is selected at the centre of the frame. Then at the encoder a ZFE is selected to represent this prototype segment. At the decoding stage the ZFE segments are interpolated, according to Section 17.6.1 to 17.6.5, in order to produce a smooth excitation waveform, which is subsequently passed through the LPC STP synthesis filter to reconstruct the original speech. The time misalignment introduced by the interpolation process described earlier can be clearly seen, where the prototype shifting is caused by the need to have an integer number of pitch prototype segments during the interpolation region. The synthesized waveform does not constitute a strict waveform replica of the original speech, which is the reason for the coder's low SEGSNR. However, it produces perceptually good speech quality.

Figure 17.12 portrays a voiced speech section, where the same process as in Figure 17.11 is followed. The synthesized waveform portrays a similar smooth waveform evolution to the input speech, but the synthesized waveform has problems maintaining the waveform's amplitude throughout all the prototype segment's resonances. McCree and Barnwell [400] suggest that this type of waveform would benefit from the postfilter described in Section 15.6. Thus far, only voiced speech frames have been discussed, hence next we provide a brief description of the unvoiced frame encoding procedure.

17.7 Unvoiced Speech

For frames that are classified as unvoiced, at the decoder a random Gaussian sequence is used as the excitation source. The same noise generator was used for the PWI-ZFE coder and the basic LPC vocoder of Chapter 15, namely the Box-Muller algorithm, which is used to produce a Gaussian random sequence scaled by the RMS energy of the LPC STP residual, where the noise generation process was described in Section 15.4.

Finally, the operation of an adaptive postfilter within the PWI-ZFE coder is examined.

17.8 Adaptive Postfilter

The adaptive postfilter from Section 15.6 was also used for the PWI-ZFE speech coder, however, the adaptive postfilter parameters were reoptimized becoming $\alpha_{pf} = 0.75$, $\beta_{pf} = 0.45$, $\mu_{pf} = 0.60$, $\gamma_{pf} = 0.50$, $g_{pf} = 0.00$ and $\xi_{pf} = 0.99$. Finally, following the adaptive postfilter the synthesized speech was passed through the pulse dispersion filter of Section 15.7.

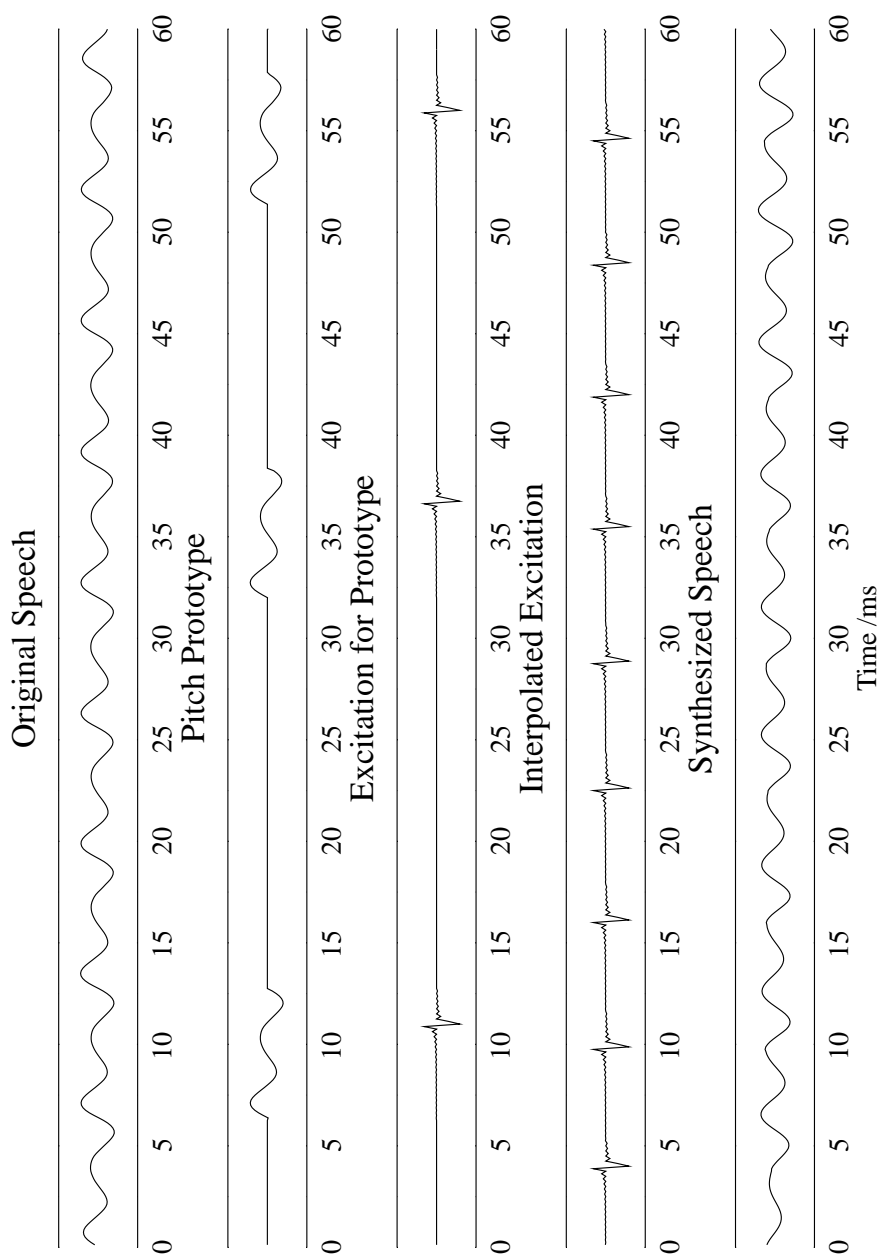


Figure 17.11: An example of the original and synthesized speech for a 60ms speech waveform from **AF2** uttering the front vowel /i/ from 'he', where the Frame Length is 20ms. The prototype segment selection and ZFE interpolation is also shown.

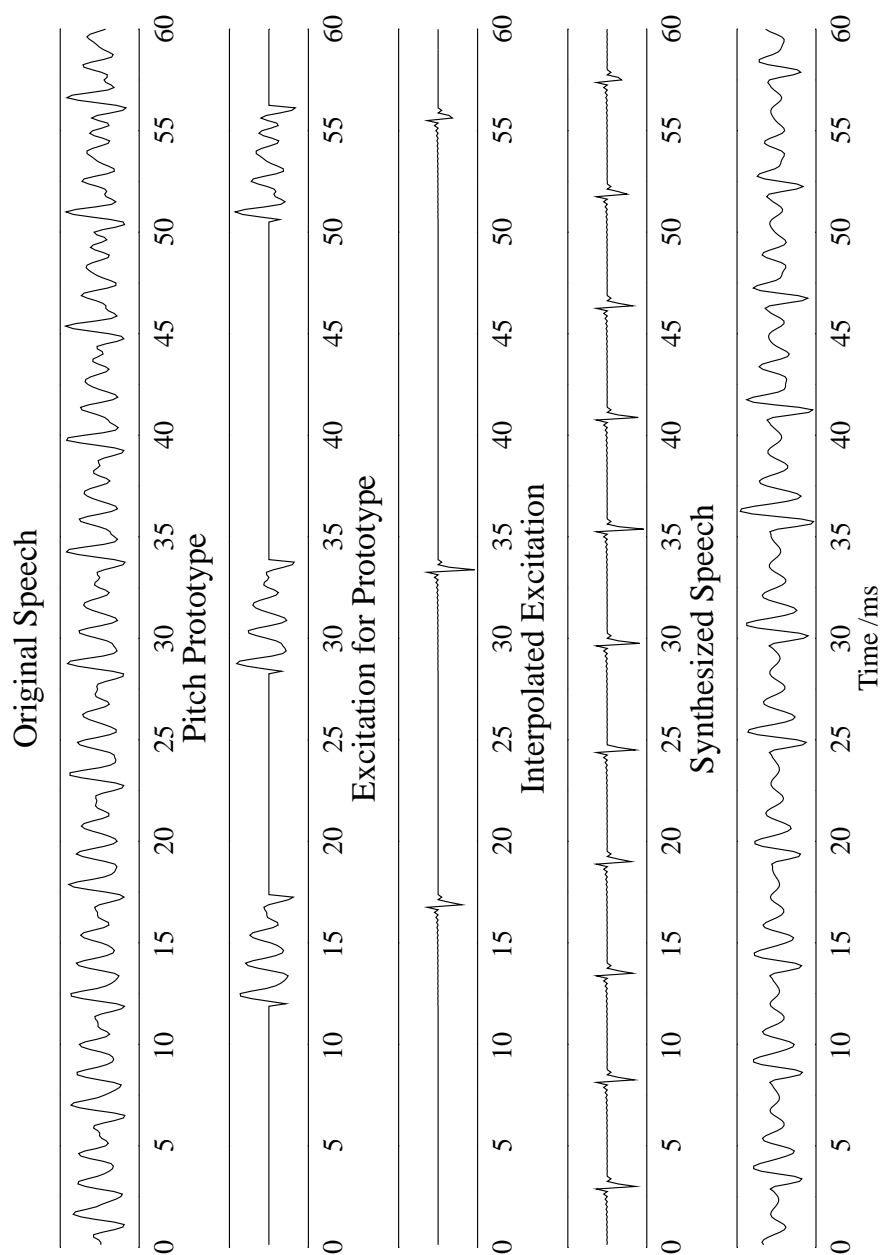


Figure 17.12: An example of three 20ms segments of the original and synthesized speech for predominantly voiced speech from **AF1 uttering the back vowel /ɔ/** 'dog'. The prototype segment selection and ZFE interpolation is also shown.

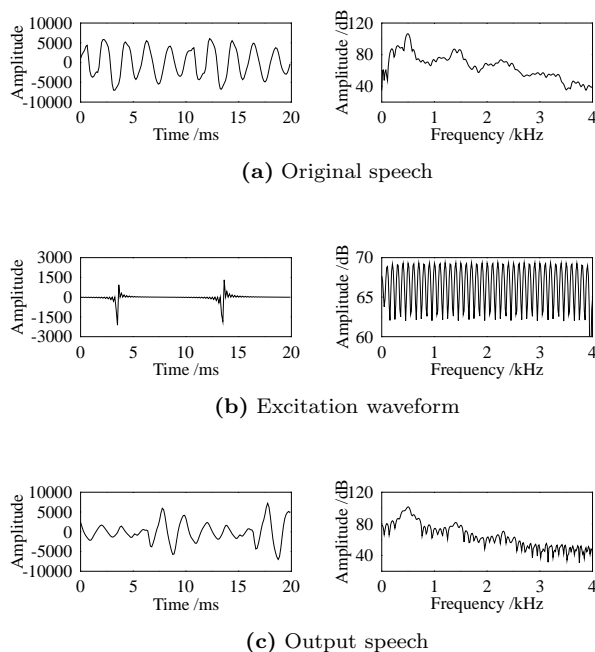


Figure 17.13: Time and frequency domain comparison of the a) original speech, b) ZFE waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the mid vowel /ɜ:/ in the utterance ‘*work*’ for the testfile **BM1**. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

Following this overview of the PWI-ZFE coder the quality of the reconstructed speech is assessed.

17.9 Results for Single Zinc Function Excitation

In this Section the performance of the PWI-ZFE speech coder described in this chapter is assessed. Figures 17.13, 17.14 and 17.15 show examples of the original and synthesized speech in the time and frequency domain for sections of voiced speech, with these graphs described in detail next. These detailed speech frames were also used to examine the LPC vocoder of Chapter 15.1, hence, Figure 17.13 can be compared to Figure 15.21, Figure 17.14 to Figure 15.22 and Figure 17.15 can be gauged against Figure 15.23.

The speech segment displayed in Figure 17.13 is a 20ms frame from testfile BM1. The reproduced speech is of similar evolution to the original speech, but cannot maintain the amplitude for the decaying resonances within each pitch period, which is due to the concentrated pulse-like nature of the ZFE. From Figure 17.13(a) and 17.13(c)

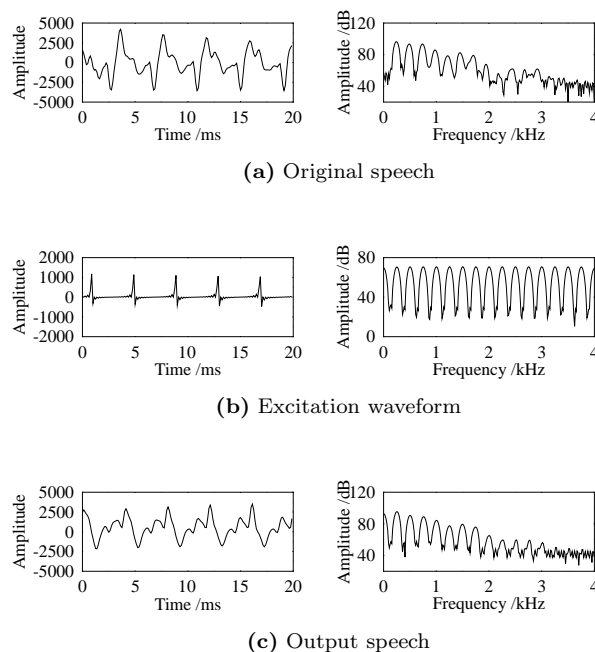


Figure 17.14: Time and frequency domain comparison of the a) original speech, b) ZFE waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the liquid /r/ in the utterance ‘rice’ for the testfile BF2. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

a time misalignment between the original and synthesized waveform is present, where the cause of the misalignment was described in Section 17.6, specifically, the interpolation region must contain an integer number of pitch prototype segments, hence, often requiring the interpolation region to be extended or shortened. Consequently, the later pitch prototype segments are shifted slightly, introducing the time misalignment seen in Figure 17.13(c). In the frequency domain the overall spectral envelope match between the original and synthesized speech is good, but as expected, the associated SEGSR is low due to the waveform misalignment experienced.

The speech segment displayed in Figure 17.14 shows the performance of the PWI-ZFE coder for the testfile BF2. Comparing Figure 17.14(c) with Figure 15.22(h), it can be seen that the synthesized waveforms in both the time and frequency domain are similar. Observing the frequency domain graphs, it is noticeable that the inclusion of unvoiced speech above 1800Hz is not modelled well by the distinct voiced-unvoiced nature of the PWI-ZFE scheme. The introduction of mixed-multiband excitation in Chapter 18 is expected to improve the representation of this signal.

The speech segment displayed in Figure 17.15 is for the testfile BM2. The synthesized speech waveform displayed in Figure 17.15(c) is noticeably better than the

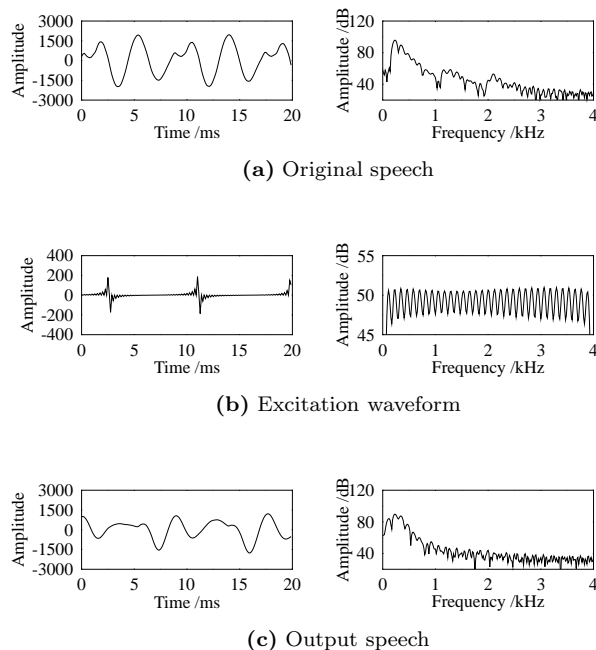


Figure 17.15: Time and frequency domain comparison of the a) original speech, b) ZFE waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the nasal /n/ in the utterance ‘**thrown**’ for the testfile **BM2**. These signals can be compared to the basic vocoder’s corresponding signals in Figure 15.23.

output speech in Figure 15.23(h). For Figure 17.15(c) the first formant is modelled well, however, the upper two formants are missing from the frequency spectrum, which is a failure in the LPC STP process and will persist in all of our developed speech coders.

Informal listening tests showed that the reproduced speech for the PWI-ZFE speech coder contained less ‘buzziness’ than the LPC vocoder of Chapter 15.

The bit allocation of the ZFE coder is summarized in Table 17.10. For unvoiced speech the RMS parameter requires the 5-bits described in Section 15.4, with the boundary shift parameter b_s offset requiring a maximum of:

$$\frac{\text{frame length}}{\text{minimum pitch}} = \frac{160}{20} = 8$$

values or 3-bits to encode.

For voiced speech the pitch period can vary from $20 \rightarrow 147$ samples, thus requiring 7-bits for transmission. Section 17.5.2 justified the use of 6-bits to SQ the A_1 and B_1 ZFE amplitude parameters.

parameter	unvoiced	voiced
LSFs	18	18
v/u flag	1	1
RMS value	5	-
b_s offset	3	-
pitch	-	7
A_1	-	6
B_1	-	6
total/20ms	27	38
bit rate	1.35kbps	1.90kbps

Table 17.10: Bit allocation table for the investigated 1.9kbps PWI-ZFE coder.

Operation /MFLOP	pitch period=20	pitch period=147
Pitch detector	2.67	2.67
ZFE minimization	0.30	11.46
Total	2.97	14.13

Table 17.11: Total maximum and minimum computational complexity for a PWI-ZFE coder.

The computational complexity of the speech coder is dominated by the ZFE minimization loop, even when using a constrained search. Table 17.11 displays the computational complexity of the coder for a pitch period of 20 samples or 147 samples.

17.10 Error Sensitivity of the 1.9kbps PWI-ZFE Coder

In this chapter we have investigated the design of a 1.9kbps speech coder employing PWI-ZFE techniques. However, we have not examined the speech coder's performance within a communications system, specifically its robustness to transmission errors. In this Section we study how the degradation caused by a typical mobile environment affects the PWI-ZFE output speech quality.

The degradation in the PWI-ZFE speech coder's performance is caused by the hostile nature of a mobile communications environment. A mobile environment typically contains both fast and slow fading, which affects the signal level at the receiver. Additionally, many different versions of the signal arrive at the receiver, each having taken different paths with different fading characteristics and different delays, thus introducing inter-symbol interference. It is these mobile environment characteristics which introduce errors into the parameters received by the speech decoder.

In this Section we commence by examining how possible errors at the decoder would affect the output speech quality and introduce some error correction techniques. These errors are then examined in terms of objective speech measures and informal listening tests. We then consider dividing the transmission bits into protection classes, which is a common technique that is adopted to afford the most error sensitive bits the greatest protection. Finally, we demonstrate the speech coder's performance for different transmission environments.

17.10.1 Parameter Sensitivity of the 1.9kbps PWI-ZFE coder

In this Section we consider the importance of the different PWI-ZFE parameters of Table 17.10 in maintaining synthesized speech quality. Additionally, we highlight checks that can be made at the decoder, which may indicate errors and suggest error correction techniques. Considering the voiced and unvoiced speech frames separately, the speech coder has 10 different parameters that can be corrupted, where the vector quantized LSFs, described in Section 15.2.2, can be considered to be four different groups of parameters. These parameters have between 7 bits, for the pitch period, and a single bit, for the voiced-unvoiced flag, which can be corrupted. In total there are 46 different bits, namely the 38 voiced bits of Table 17.10 and the RMS and b_s unvoiced parameters.

Finally, we note that due to the interpolative nature of the PWI-ZFE speech coder any errors that occur in the decoded bits will affect more than just the frame where the error occurred.

17.10.1.1 Line Spectrum Frequencies

The LSF vector quantizer, described in Section 15.2.2 and taken from G.729 [269], represents the LSF values using four different parameters. The LSF VQ consists of a 4th order moving average (MA) predictor, which can be switched on or off with the flag L0. The vector quantization is then performed in two stages. A 7-bit VQ index, L1, is used for the first stage. The second stage VQ is a split vector quantizer, using the indices L2 and L3, with each codebook containing 5-bits.

17.10.1.2 Voiced-Unvoiced Flag

It is anticipated that the voiced-unvoiced flag will be the most critical bit for the successful operation of the PWI-ZFE speech coder. The very different excitation models employed for voiced and unvoiced speech mean that if the wrong type of excitation is adopted, this is expected to have a serious degrading effect.

At the decoder it is possible to detect isolated errors in the voiced-unvoiced flag, namely V-U-V and U-V-U sequences in the $N+1$, N , $N-1$ frames. These sequences will indicate an error, since at the encoder they were prohibited frame combinations, as described in Section 17.2.1. However, the PWI-ZFE decoder does not operate on a frame-by-frame basis, instead it performs interpolation between the prototype segments of frame N and $N + 1$, as described in Section 17.2.1. Thus, without introducing an extra 20ms delay, by performing the interpolation between frames $N - 1$ and N , it is impossible to completely correct an isolated error in the voiced-unvoiced flag.

17.10.1.3 Pitch Period

The pitch period parameter of Table 17.10 is only sent for voiced frames, where having the correct pitch period is imperative for producing good quality synthesized speech. In Section 16.5.2 some simple pitch period correction was already performed, where checks were made to ensure a smooth pitch track is followed. By repeating

this pitch period correction at the decoder the effect of an isolated pitch period error can be reduced. However, similarly to the voiced-unvoiced flag, the use of frames N and $N + 1$ in the interpolation process permits an isolated pitch period to have a degrading effect.

17.10.1.4 Excitation Amplitude Parameters

The ZFE amplitude parameters, A and B, control the shape of the voiced excitation. The A and B parameters of Table 17.10 can have both positive and negative values, however, as described in Section 17.3.4, the phase of the amplitude parameters must be maintained throughout the voiced sequence. At the decoder it is possible to maintain phase continuity for the amplitude parameter in the presence of an isolated error, with the correction that if the phase of the A or B parameter has been found to change during a voiced sequence, then the previous A or B parameter can be repeated.

17.10.1.5 Root Mean Square Energy Parameter

For unvoiced speech frames the excitation is formed from random Gaussian noise scaled by the received RMS energy value, seen in Table 17.10 and as described in Section 17.7. Thus, if corruption of the RMS energy parameter occurs, then the energy level of the unvoiced speech will be incorrect. However, since the speech sound is a low pass filtered, slowly varying process, abrupt RMS changes due to channel errors can be detected and mitigated.

17.10.1.6 Boundary Shift Parameter

The boundary shift parameter, b_s , of Table 17.10 is only sent for unvoiced frames and defines the location where unvoiced speech becomes voiced speech, or vice versa. The corruption of the boundary shift parameter will move this transition point, an event which is not amenable to straight forward error concealment.

17.10.2 Degradation from Bit Corruption

Following this discussion on the importance of the various PWI-ZFE parameters and the possible error corrections which could be performed at the speech decoder, we now investigate the extent of the degradation which errors cause to the reproduced speech quality. The error sensitivity is examined by separately corrupting each of the 46 different voiced and unvoiced bits of Table 17.10, where 18 LSF plus the v/uv bits are sent for all frames, additionally 19 bits are only sent for voiced frames and 8 bits are only sent for unvoiced frames. For each selected bit the corruption was inflicted 10% of the time. Corrupting a bit for 10% of the time is a compromise between consistently or constantly corrupting the bit in all frames and corrupting the bit in only a single isolated frame. If the bit is constantly corrupted then any error propagation effect is masked, while corrupting the bit in only a single frame requires that for completeness every possible frame is taken to be that single frame, resulting in an arduous process.

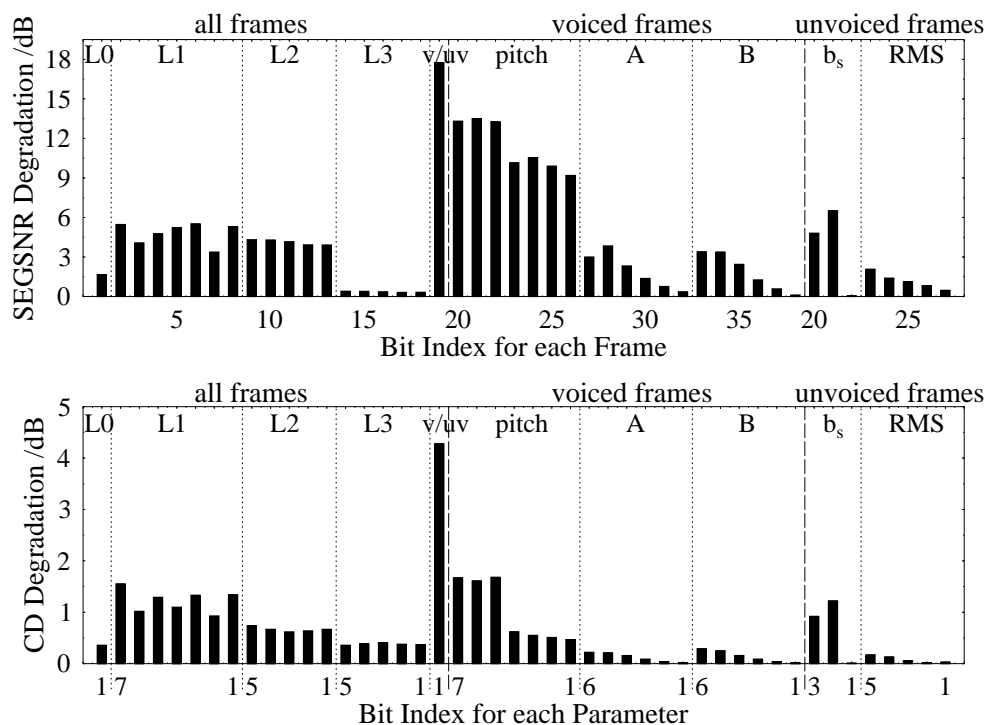


Figure 17.16: The error sensitivity of the different transmission bits for the 1.9kbps PWI-ZFE speech coder. The graph is divided into bits sent for all speech frames, bits sent only for voiced frames and bits sent only for unvoiced frames. For the CD degradation graph containing the bit index for each parameter bit 1 is the least significant bit.

Figure 17.16 displays the averaged results for the speech files AM1, AM2, AF1, AF2, BM1, BM2, BF1, BF2, described in Section 14.4. The SEGSNR and CD objective speech measures, described in Section 14.3.1, were used to evaluate the degradation effect. Additionally, the synthesized corrupted speech due to the different bit errors was compared through informal listening tests.

Observing Figure 17.16 it can be seen that both the SEGSNR and CD objective measures rate the error sensitivity of the different bits similarly, both indicating that the voiced-unvoiced flag being correct is the most critical for successful synthesis of the output speech. This was confirmed by listening to the synthesized speech, which was frequently unintelligible when there was 10% error in the voiced-unvoiced flag bit. Additionally, from Figure 17.16 it can be seen that both the pitch period and boundary shift parameters produce a significant degradation due to bit errors. However, informal listening tests do not indicate such significant quality degradation, although an incorrect pitch period does produce audible distortion. It is suggested that the time misalignment introduced by the pitch period and boundary shift parameter errors is

Classes	Coding Bits						
1	v/uv						
2	L1[7] pitch[7] A[6]	L1[5] pitch[6] A[5]	L1[3] pitch[5] B[6]	L1[1] pitch[4] B[5]	pitch[3]	pitch[2]	pitch[1]
3	L0 L2[5] L3[5] A[4] B[4]	L1[6] L2[4] L3[4] A[3] B[3]	L1[4] L2[3] L3[3] A[2] B[2]	L1[2] L2[2] L3[2] A[1] B[1]	L2[1] L3[1]		

Table 17.12: The transmission classes for the bits of the 1.9kbps PWI-ZFE speech coder, with class 1 containing the most error sensitive bits and class 3 bits requiring little error protection.

artificially increasing the SEGSNR and CD degradation values.

Thus, while the SEGSNR and CD objective measures indicate the relative sensitivities of the bits within each parameter, more accurate interpretation of the sensitivity of each parameter has to rely more on informal listening tests.

17.10.2.1 Error Sensitivity Classes

The SEGSNR and CD objective measures together with the informal listening tests allow the bits to be grouped into three classes for transmission to the decoder. These classes are detailed in Table 17.12, where class 1 requires the greatest protection and class 3 requires the least protection.

In Table 17.12 the error sensitivity classes are based on the bits sent every speech frame and bits sent only for voiced frames, giving 38 bits. For unvoiced frames the boundary parameter shift, b_s , is given the same protection as the most significant three pitch period bits, while the RMS value is given the same protection as the least significant four pitch period bits and A[6].

Class 1 contains only the voiced-unvoiced flag, which has been identified as being very error sensitive. Class 2 contains 15 bits, while class 3 contains 22 bits.

The relative bit error sensitivities have been used to improve channel coding within a GSM-like speech transceiver [443] and a FRAMES-like speech CDMA transceiver [444].

Following this analysis of the performance of a PWI-ZFE speech coder, using a single ZFE to represent the excitation, the potential for speech quality improvement with extra ZFE pulses is examined.

17.11 Multiple Zinc Function Excitation

So far in this chapter a single ZFE pulse has been employed to represent the voiced excitation. However, a better speech quality may be achieved by introducing more ZFE pulses [411]. The introduction of extra ZFEs will be at the expense of a higher

$Total K$	rescaled ZFE needed	> 1 ZFE rescaled	> 2 ZFE rescaled	> 3 ZFE rescaled	> 4 ZFE rescaled
1	12.9%	-	-	-	-
2	20.3%	5.1%	-	-	-
3	33.0%	7.1%	1.0%	-	-
4	42.3%	16.5%	5.1%	0.6%	-
5	57.9%	28.4%	10.7%	2.5%	0.1%

Table 17.13: The percentage of speech frames requiring previous ZFEs to be scaled and repeated, for K ZFE pulses in PWI-ZFE coders.

bit rate, thus, a dual-mode PWI-ZFE speech coder could be introduced to exploit an improved speech quality when traffic density of the system permits.

Revisiting the ZFE error minimization process of Section 17.3.1, where due to the orthogonality of the zinc basis functions the weighted error signal upon using k ZFE pulses is given by:

$$E_w^{k+1} = \sum_{n=1}^P (e_w^{k+1}(n))^2 \quad (17.37)$$

where P is the length of the prototype segment, over which minimization is carried out, with the synthesized weighted speech represented by:

$$\bar{s}_w(n) = \sum_{k=1}^K z_k(n) * h(n) \quad (17.38)$$

where $z_k(n)$ is the k^{th} ZFE pulse, K is the number of pulses being employed and $h(n)$ is the impulse response of the weighted LPC synthesis filter.

17.11.1 Encoding Algorithm

The encoding process for a single ZFE was previously described in Table 17.1 and Figure 17.2. For a multiple ZFE arrangement the same process is followed, but the number of ZFE pulses is extended to K , as shown in Figure 17.17 and described next. Thus, for the phase constrained frame, which we also refer to as the phase restriction frame, a phase is determined independently for each of the K excitation pulses. Similarly, for other voiced frames the phase of the k^{th} pulse is based on the phase restriction for the k^{th} pulses. Furthermore, if a suitable ZFE is not found for the k^{th} ZFE pulse in frame N , then the k^{th} ZFE in frame $N - 1$ is scaled and reused.

For scenarios with a different number ZFE pulse per prototype segment Table 17.13 displays the percentage of voiced frames, where some scaling from the previous frame's ZFE pulses must be performed. It can be seen that with 3 ZFE pulses employed 1/3 of the voiced frames contain scaled ZFE pulses from the previous frame. Additionally, some frames have several scaled ZFE pulses from the previous frame.

The implementation of the single ZFE, in Section 17.3.3, showed that for smooth interpolation it is beneficial to constrain the locations of the ZFE pulses. Constraining the K ZFE locations follows the same principles as those used in determining the

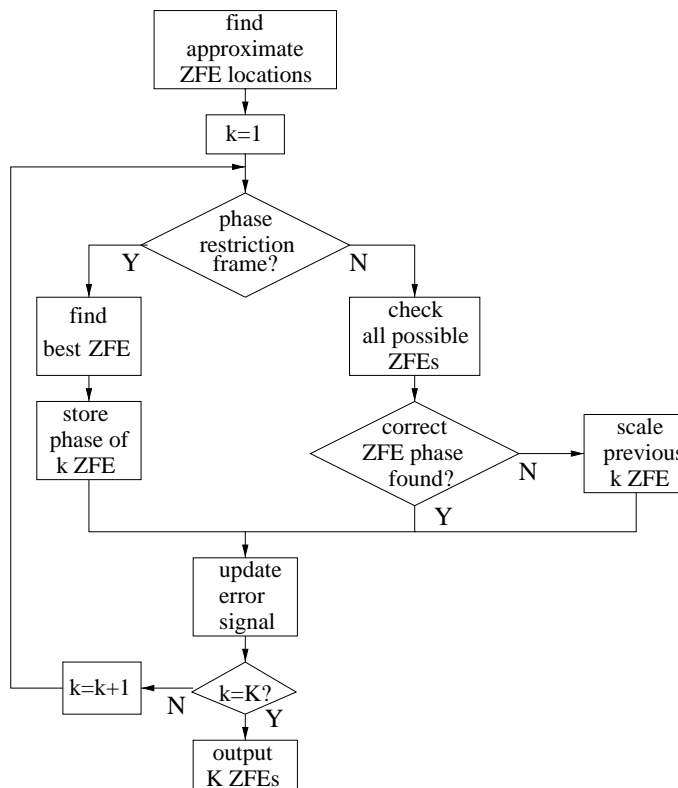


Figure 17.17: The control structure for selecting multiple ZFEs in PWI-ZFE coders.

single ZFE location, but it was extended to find K constrained positions. For the first voiced frame the largest K impulses, determined by wavelet analysis according to Chapter 16 and located within the prototype segment, are selected for the positions the ZFE pulses must be in proximity to. For further voiced frames the impulses from the wavelet analysis are examined, with the largest impulses near the K ZFE pulses in frame $N - 1$ selected as excitation. If no impulse is found near the k^{th} ZFE location in frame $N - 1$, this position is repeated as the k^{th} ZFE in frame N . It is feasible that there will be less than K wavelet analysis impulses within the prototype segment, thus in this situation the extra ZFEs are set to zero. They are subsequently introduced, when impulses occur within the prototype segment that are unrelated to any ZFE pulses in frame $N - 1$.

The SEGSR values achieved for the minimization process at the encoder with different numbers of ZFE pulses per prototype segment indicate the excitation representation improvement. Figure 17.18 displays the results, showing that the improvement achieved by adding extra pulses saturates as the number of ZFE pulses increases, so when eight ZFE pulses are employed no further SEGSR gain is achieved. The limit in SEGSR improvement is due to the constraint that ZFE pulses are expected to

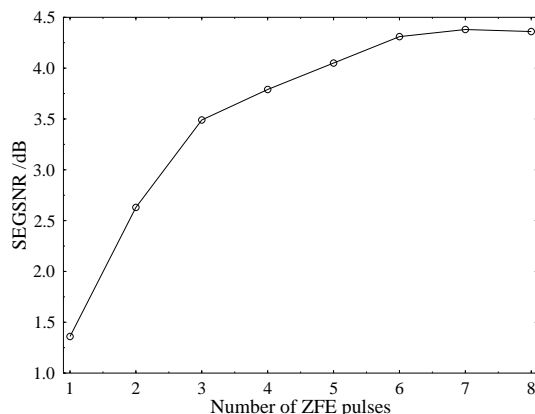


Figure 17.18: The SEGSNR achieved at the encoder minimization process for different number of ZFE pulses used in the representation. The inclusion of each new ZFE pulse requires 19 extra bits/20ms, or 0.95kbps extra bit rate, for the encoding of the A_k and B_k parameters and the additional ZFE pulse positions λ_k , as seen in Table 17.14.

be near the instants of glottal closure found by the wavelet analysis. There will be a limited number of impulses within the prototype segment, thus a limited number of ZFE pulses can be employed for each prototype segment. The performance of a three pulse ZFE scheme at the encoder is given in Figure 17.19, which can be compared with the performance achieved by a single ZFE, shown in Figure 17.9. It can be seen that the addition of two extra ZFE pulses improves the excitation representation, particularly away from the main resonance.

At the decoder the same interpolation process implemented for the single ZFE is employed, as described in Section 17.6, again extended to K ZFE pulses. For all ZFE pulses the amplitude parameters are linearly interpolated, with the ZFE pulse position parameter and prototype segment location assumed at the decoder, as in the single pulse coder of earlier sections. Explicitly, the k^{th} ZFE pulse position parameter is kept at the same location within each prototype segment. For the three pulse PWI-ZFE scheme the adaptive postfilter parameters were reoptimized becoming $\alpha_{pf} = 0.75$, $\beta_{pf} = 0.45$, $\mu_{pf} = 0.40$, $\gamma_{pf} = 0.50$, $g_{pf} = 0.00$ and $\xi_{pf} = 0.99$.

17.11.2 Performance of Multiple Zinc Function Excitation

A three pulse ZFE scheme was implemented to investigate the potential for improved speech quality using extra ZFE pulses. Three excitation pulses were adopted to study the feasibility of a speech coder at 3.8kbps, where the bit allocation scheme was given in Table 17.14.

Figure 17.20 displays the performance of a three pulse ZFE scheme for the mid vowel /ɜ/ in the utterance 'work' for the testfile BM1. The identical portion of speech synthesized using a single ZFE was given in Figure 17.13. From Figure 17.20(b) it can be seen that the second largest ZFE pulse is approximately half-way between the

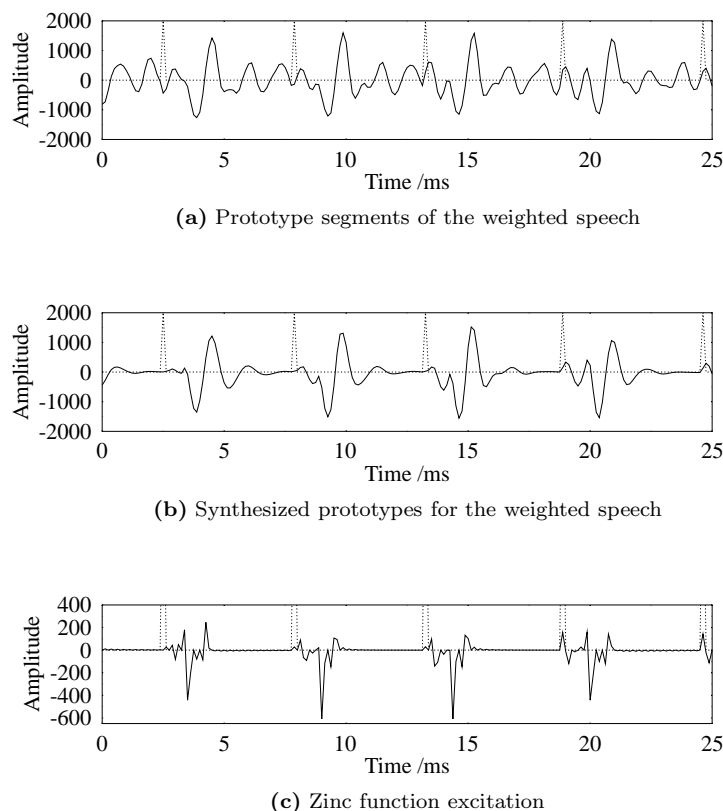


Figure 17.19: Demonstrating the process of analysis-by-synthesis encoding for prototype segments that have been concatenated to produce a smoothly evolving waveform, with the excitation represented by three ZFE pulses. The dotted lines in the Figure indicate the boundaries between prototype segments.

largest ZFE pulses. In the frequency spectrum the pitch appears to be 200Hz, which is double the pitch from Figure 17.13(b). The pitch doubling is clearly visible in the time and frequency domain of Figure 17.20(c). For this speech frame the addition of extra ZFE pulses fails to improve the speech quality, where this is due to the secondary excitation pulse producing a pitch doubling effect in the output speech.

Figure 17.21 displays the results from applying a three pulse ZFE scheme to a 20ms frame of speech from the testfile BF2. The same speech frame was investigated in Figures 17.14 and 15.22. Observing Figure 17.21(b) it can be seen that similarly to Figure 17.20(b) a ZFE pulse is placed midway between the other ZFE pulses, however, since this pulse has much less energy it does not have a pitch doubling effect. When compared with the single ZFE of Figure 17.15(c) the multiple ZFEs combine to produce a speech waveform, shown in Figure 17.21(c), much closer in both the time and frequency domain to the original, although at the cost of a higher

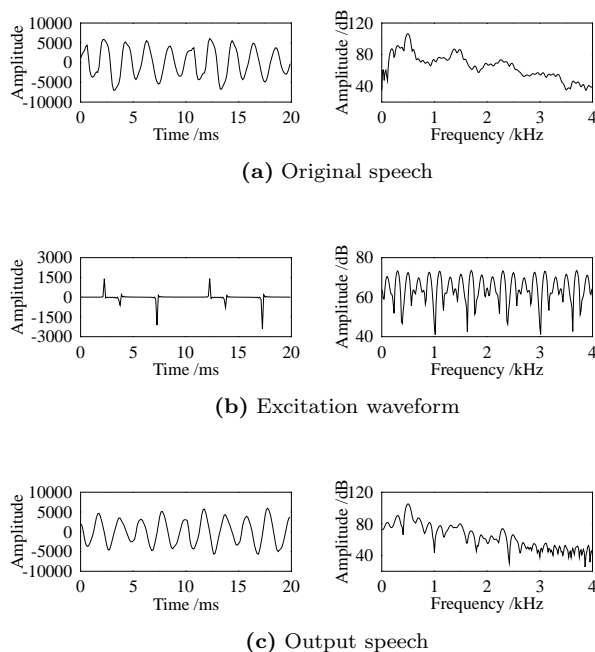


Figure 17.20: Time and frequency domain comparison of the a) original speech, b) Three pulse ZFE waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the mid vowel /ɜ:/ in the utterance ‘work’ for the testfile BM1. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

bit rate and complexity.

Figure 17.22 portrays a three pulse ZFE scheme applied to a speech frame from the testfile BM2, which can be compared with Figure 17.15. From Figure 17.22(b) it can be seen that no pitch doubling occurs. For this speech frame the limiting factor in reproducing the original speech are the missing formants. However, observing Figure 17.22(c) demonstrates that three ZFE pulses results in an improved performance compared with a single ZFE.

Informal listening tests were conducted using the PWI-ZFE speech coder with three ZFE pulses, where it was found that sudden and disconcerting changes could occur in the quality of the reproduced speech. It is suggested that this effect was created by the varying success of the excitation to represent the speech. Additionally, for many speech files there was a background roughness to the synthesized speech. The problems with implementing a multiple ZFE pulse scheme are caused by the interpolative nature of the speech coder. The benefits, which are gained in improved representation of the excitation signal, are counteracted by increased problems in both obeying phase restrictions and in creating a smoothly interpolated synthesized speech waveform.

For the 3.8kbps multiple ZFE speech coder the extra bits are consumed by the two

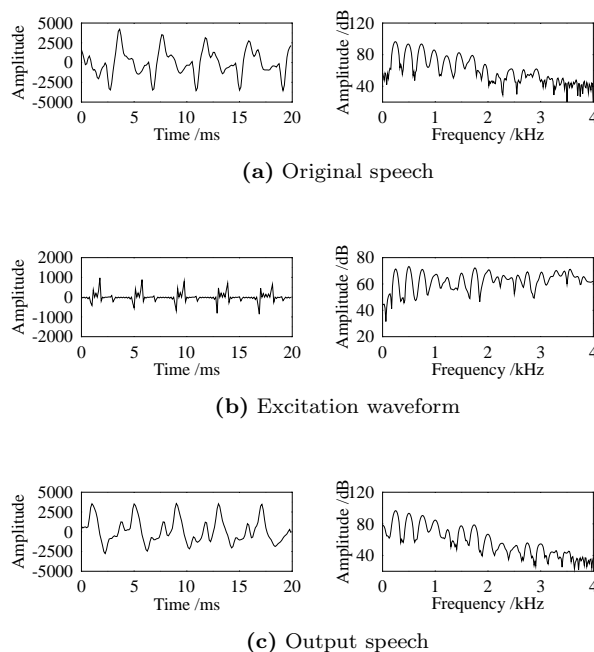


Figure 17.21: Time and frequency domain comparison of the a) original speech, b) Three pulse ZFE waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the liquid /r/ in the utterance ‘rice’ for the testfile BF2. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

extra ZFE pulses, with the bit allocation detailed in Table 17.14. The location of the two extra ZFE pulses, λ_2 and λ_3 , with respect to the first ZFE pulse, must be transmitted to the decoder, while, similarly to the single ZFE coder, the first pulse location can be assumed at the decoder. With a permissible pitch period range of $20 \rightarrow 147$ samples 7 bits are required to encode each position parameter, λ . This parameter only requires transmission for the first frame of a voiced sequence, since for further frames the pulses are kept in the same location within the prototype region, as it was argued in Section 17.6.3. The A and B amplitude parameter for the extra ZFE pulses are scalar quantized to 6-bits.

In order to produce a dual-rate speech coder it must be possible to change the coder’s transmission rate during operation. In this multiple ZFE scheme, if a ZFE pulse were omitted from the frame, reducing the bit rate, at the decoder the ZFE pulse would be interpolated across the interpolation region to zero. Similarly, if an extra ZFE pulse was harnessed, then at the decoder the ZFE would be interpolated from zero. This interpolation from zero degrades the assumption that the previous prototype segment at the encoder is similar to the previous interpolation region at the decoder. Thus it is prudent to only permit coding rate changes between voiced

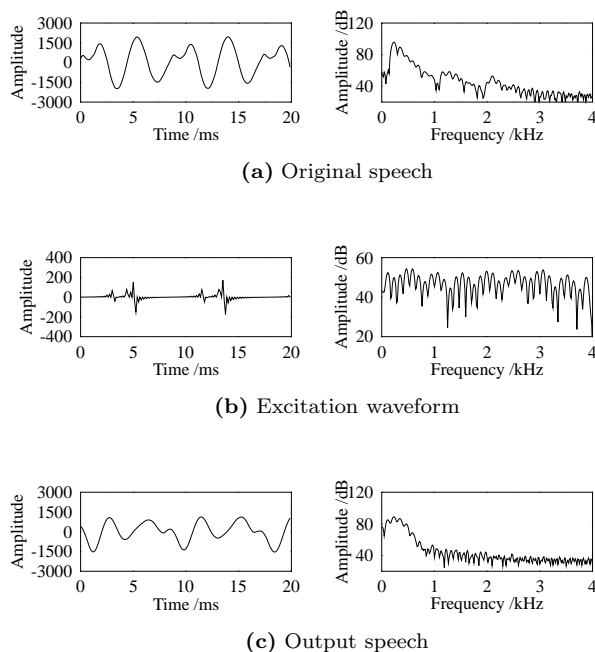


Figure 17.22: Time and frequency domain comparison of the a) original speech, b) Three pulse ZFE waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the nasal /n/ in the utterance ‘*thrown*’ for the testfile BM2. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

frame sequences.

17.12 A Sixth-rate, 3.8 kbps GSM-like Speech Transceiver¹

17.12.1 Motivation

Although the standardisation of the third-generation wireless systems has been completed, it is worthwhile considering potential evolutionary paths for the mature GSM system. This tendency was hallmarked by the various GSM Phase2 proposals, endeavouring to improve the services supported or by the development of the half-rate and enhanced full-rate speech codecs. In this section two potential improvements and their interactions in a source-sensitivity matched transceiver are considered, namely employing an approximately sixth-rate, 1.9 kbps speech codec and turbo coding [157,325]

¹This section is based on F.C.A Brooks, B.L Yeap, J.P Woodard and L. Hanzo: A Sixth-rate, 3.8 kbps GSM-like Speech Transceiver, ACTS’98, Rhodes, Greece

parameter	voiced
LSFs	18
v/u flag	1
pitch	7
1st pulse	
A_1	6
B_1	6
2nd pulse	
λ_2	7
A_2	6
B_2	6
3rd pulse	
λ_3	7
A_3	6
B_3	6
total/20ms	76
bit rate	3.80kbps

Table 17.14: Bit allocation table for voiced speech frames in the 3.8kbps investigated PWI-ZFE coder employing three ZFEs.

in conjunction with the GSM system's Gaussian Minimum Shift Keying (GMSK) partial response modem.

The bitallocation of the 1.9 kbps PWI-ZFE speech codec was summarised in Table 17.10, while its error sensitivity was quantified in Section 17.10. The SEGSNR and CD objective measures together with the informal listening tests allow the bits to be ordered in terms of their error sensitivities. The most sensitive bit is the voiced-unvoiced flag. For voiced frames the three most significant bits (MSB) in the LTP delay are the next most sensitive bits, followed by the four least significant LTP delay bits. For unvoiced frames the boundary parameter shift, j , is given the same protection as the most significant three pitch period bits, while the RMS value is given the same protection as the group of four least significant pitch period bits and bit $A[6]$, the LSB of the ZFE amplitude A .

17.12.2 The Turbo-coded Sixth-rate 3.8 kbps GSM-like System

The amalgamated GSM-like system [34] is illustrated in Figure 17.23. In this system, the 1.9kbps speech coded bits are channel encoded with a $\frac{1}{2}$ rate convolutional or turbo encoder [157,325] with an interleaving frame-length of 81 bits, including termination bits. Therefore, assuming negligible processing delay, 162 bits will be released every 40ms, or two 20ms speech frames, since the 9x9 turbo-interleaver matrix employed requires two 20ms, 38 bit, speech frames before channel encoding commences. Hence we set the data burst length to be 162 bits. The channel encoded speech bits are then passed to a channel interleaver. Subsequently, the interleaved bits are modulated using Gaussian Minimum Shift Keying (GMSK) [34] with a normalised bandwidth,

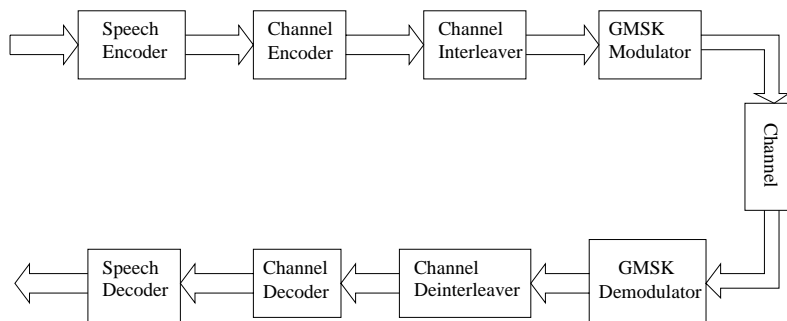


Figure 17.23: GSM-like system block diagram

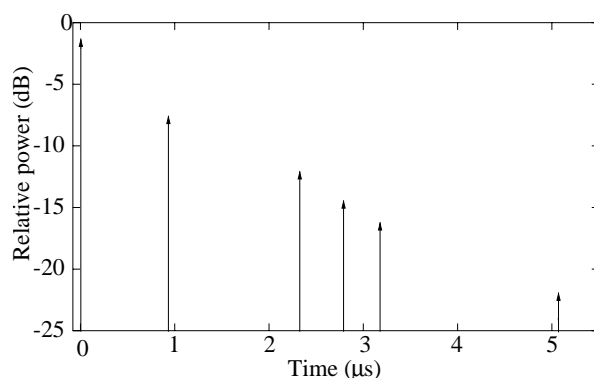


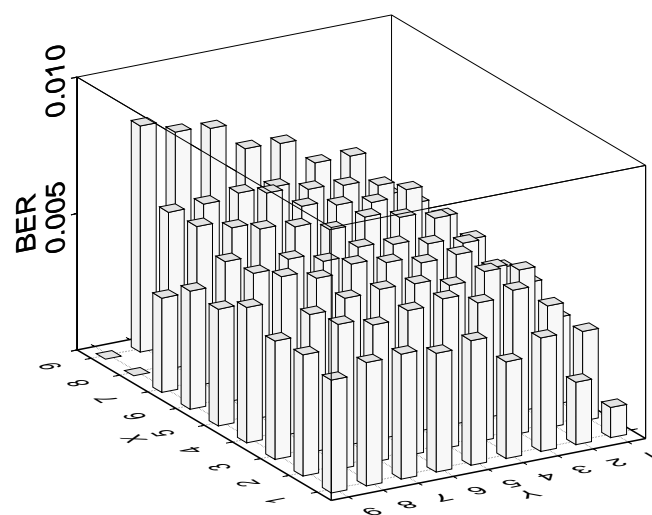
Figure 17.24: The impulse response of the COST207 Typical Urban channel used [390]

$B_n = 0.3$ and transmitted at 271Kbit/s across the COST 207 [390] Typical Urban channel model. Figure 17.24 is the Typical Urban channel model used and each path is fading independently with Rayleigh statistics, for a vehicular speed of 50km/h or 13.89 ms^{-1} and transmission frequency of 900 MHz.

The GMSK demodulator equalises the received signal, which has been degraded by the wideband fading channel, using perfect channel estimation [34]. Subsequently, soft outputs from the demodulator are deinterleaved and passed to the channel decoder. Finally, the decoded bits are directed towards the speech decoder in order to extract the original speech information. In the following sub-sections, the channel coder and interleaver/deinterleaver, and GMSK transceiver are described.

17.12.3 Turbo Channel Coding

We compare two channel coding schemes, constraint-length $K = 5$ convolutional coding as used in the GSM [34] system, and a turbo channel codec [157, 325]. The turbo codec uses two $K = 3$ so-called Recursive Systematic Convolutional (RSC) component codes employing octally represented generator polynomials of 7 and 5,



GMT Apr 29 17:47 413

Figure 17.25: The error sensitivity of the different information bits within the 9x9 block interleaver used in the turbo codec

as well as 8 iterations of the Log-MAP [445] decoding algorithm. This makes it approximately 10 times more complex than the convolutional codec.

It is well known that turbo codes perform best for long interleavers. However due to the low bit rate of the speech codec we are constrained to using a low frame length in the channel codecs. A frame length of 81 bits is used, with a 9x9 block interleaver within the turbo codec. This allows two sets of 38 coded bits from the speech codec and two termination bits to be used. The BERs of the 79 transmitted bits with the 9x9 block interleaver used for the turbo codec, for a simple AWGN channel at an SNR of 2 dB, is shown in Figure 17.25. It can be seen that bits near the bottom right hand corner of the interleaver are better protected than bits in other positions in the interleaver. By placing the more sensitive speech bits here we are able to give significantly more protection to the V/U flag and to some of the other sensitive speech bits, than to the low-sensitivity bits of Figure 17.16. Our current work investigates providing more significant un-equal error protection using turbo-codes with irregular parity bit puncturing. Lastly, an interburst channel interleaver is used, in order to disperse the bursty channel errors and to assist the channel decoders, as proposed for GSM [34].

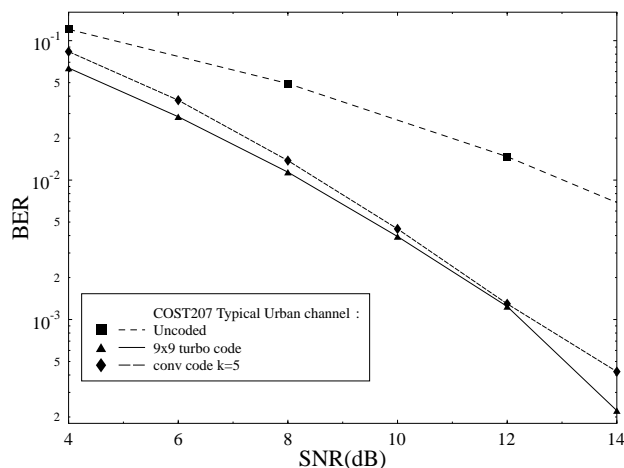


Figure 17.26: The BER performance for the turbo and convolutional coded systems over the COST 207 Typical Urban channel [390]

17.12.4 The Turbo-coded GMSK Transceiver

As mentioned in Section 17.12.2, a GMSK modulator, with $B_n = 0.3$, which is employed in the current GSM [34] mobile radio standard, is used in our system. GMSK belongs to a class of Continuous Phase Modulation (CPM) [34], and possesses high spectral efficiency and constant signal envelope, hence allowing the use of non-linear power efficient class-C amplifiers. However, the spectral compactness is achieved at the expense of Controlled Intersymbol Interference (CISI), and therefore an equaliser, typically a Viterbi Equaliser, is needed. The conventional Viterbi Equaliser (VE) [34] performs Maximum Likelihood Sequence Estimation by observing the development of the accumulated metrics, which are evaluated recursively, over several bit intervals. The length of the observation interval depends on the complexity afforded. Hard decisions are then released at the end of the equalisation process. However, since Log Likelihood Ratios (LLRs) [446] are required by the turbo decoders, we could use a variety of soft output algorithms instead of the VE, such as the Maximum A Posteriori (MAP) [329] algorithm, the Log-MAP [445], the Max-Log-MAP [447, 448], and the Soft Output Viterbi Algorithm (SOVA) [27, 449, 450]. We chose to use the Log-MAP algorithm as it gave the optimal performance, like the MAP algorithm, but at a much lower complexity. Other schemes like the Max-Log-MAP and SOVA, are computationally less intensive, but provide sub-optimal performance. Therefore, for our work, we have opted for the Log-MAP algorithm in order to obtain the optimal performance, hence giving the upper bound performance of the system.

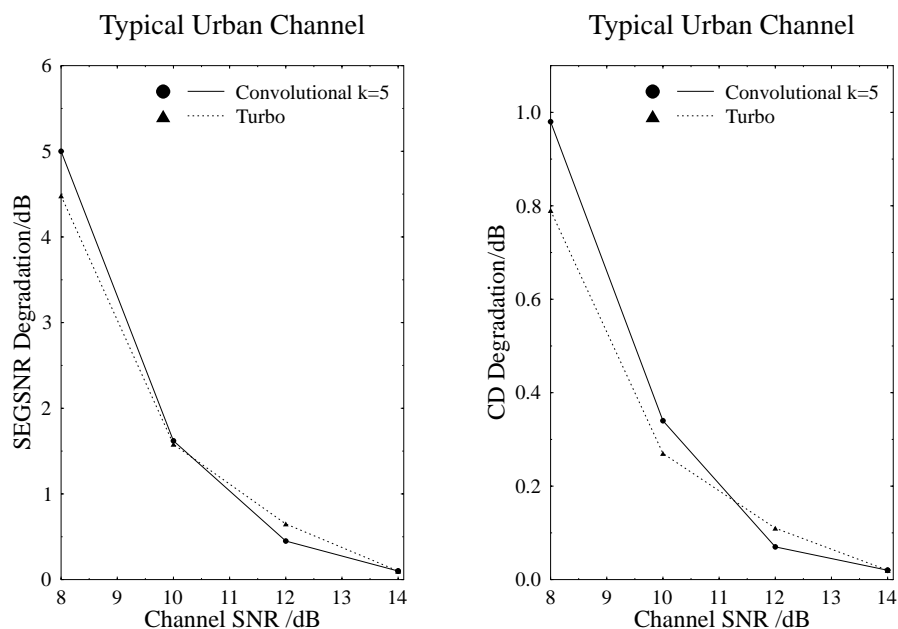


Figure 17.27: The speech degradation performance for the turbo and convolutional coded systems over the COST 207 Typical Urban channel [390]

17.12.5 System Performance Results

The performance of our sixth-rate GSM-like system was compared with an equivalent conventional GSM system using convolutional codes instead of turbo codes. The $\frac{1}{2}$ rate convolutional code [34] has the same code specifications as in the standard GSM system [34]. Figure 17.26 illustrates the BER performance over a Rayleigh fading COST207 Typical Urban channel [390] and Figure 17.27 shows the speech degradation, in terms of both the Cepstral Distance (CD) and the Segmental SNR, for the same channel. Due to the short interleaver frame length of the turbo code the turbo- and convolutionally coded performances are fairly similar in terms of both BER and speech degradation, hence the investment of the higher complexity turbo codec is not justifiable, demonstrating an important limitation of short-latency interactive turbo-coded systems. However, we expect to see higher gains for higher bit rate speech codecs, such as for example the 260bit/20ms full-rate and the enhanced full-rate GSM speech codecs, which would allow us to use larger frame lengths for the turbo code, an issue currently investigated.

17.13 Summary and Conclusions

This chapter has described a PWI-ZFE coder previously suggested by Hiotakakos and Xydeas [410]. However, their work was further developed in this chapter to reduce the bit rate and complexity, while improving speech quality. Section 17.2 to 17.4 gave an overview of the speech coder, with Figure 17.4 demonstrating the prohibitive complexity of the original ZFE optimization process proposed by Hiotakakos and Xydeas [410]. This prohibitive complexity was significantly reduced by introducing wavelets into the optimization process. Section 17.5 described the voiced speech encoding procedure, involving ZFE optimization and ZFE amplitude coefficient quantization. Energy scaling was also proposed to ensure that the original speech amplitude was maintained in the synthesized speech. The interpolation performed at the decoder was detailed in Section 17.6, where the justifications for not sending either the starting location of the prototype segment, or the ZFE position parameter, were given. The PWI-ZFE description was completed in Sections 17.7 and 17.8 which briefly described the unvoiced speech and adaptive postfilter requirements, respectively.

The PWI-ZFE speech coder at 1.9kbps was found to produce speech with a more natural quality than the basic LPC vocoder of Chapter 15. It has also been shown in this chapter that numerous benefits were attainable in reducing the computational complexity through the use of the wavelet transform of Chapter 16 with no discernible reduction in speech quality. Particularly useful was the ability of the wavelet transform to suggest instants of glottal closure. The chapter also outlined an interpolation method at the decoder which permitted the ZFE amplitude parameters to be transmitted without the position parameter, reducing the bit rate. Finally, in Section 17.11 multiple ZFE was considered, however, the quality of the synthesized speech was often found to be variable.

Chapter 18

Mixed-Multiband Excitation

18.1 Introduction

This chapter investigates the speech coding technique of Mixed-Multiband Excitation (MMBE) [225] which is frequently adopted in very low bit rate voice compression. The principle behind MMBE is that low bit rate speech coders, which follow the classical vocoder principle of Atal and Hanauer [395] invoking distinct separation into voiced-unvoiced segments, usually result in speech of a synthetic quality due to a distortion generally termed ‘buzziness’. This ‘buzzy’ quality is particularly apparent in portions of speech which contain only voiced excitation in some frequency regions, but dominant noise in other frequency bands of the speech spectrum. A classic example is the fricative class of phonemes, which contain both periodic and noise excitation sources. In low bit rate speech coders this type of speech waveform can be modelled successfully by combining voiced and unvoiced speech sources. Figure 18.1 shows the case of the voiced fricative /z/ as in ‘zoo’, which consists of voiced speech up to 1kHz and predominantly noisy speech above this frequency. Improved voiced excitation sources, such as the ZFE described in Chapter 17, can remove some of the synthetic quality of the reconstructed speech. However, the ZFE does nothing to combat the inherent problem of ‘buzziness’, which is associated with a mixed voiced-unvoiced spectrum that often occurs in human speech production.

MMBE addresses the problem of ‘buzziness’ directly through splitting the speech into several frequency bands, similarly to subband coding [352] on a frame-by-frame adapted basis. These frequency bands have their voicing assessed individually with an excitation source of pulses, noise or a mixture of both being selected for each frequency band. Figure 18.2 shows the PDF of the voicing strength for the training speech database of Table 14.1, where the voicing strength is defined later in Equation 18.11. It demonstrates that although the voicing strengths have significant peaks near the values of 0.3 and 1, representing unvoiced and voiced frames, respectively, there are a number of frames with intermediate voicing strength. It is these frames, constituting about 35% having voicing strengths between 0.4 and 0.85, which will benefit from

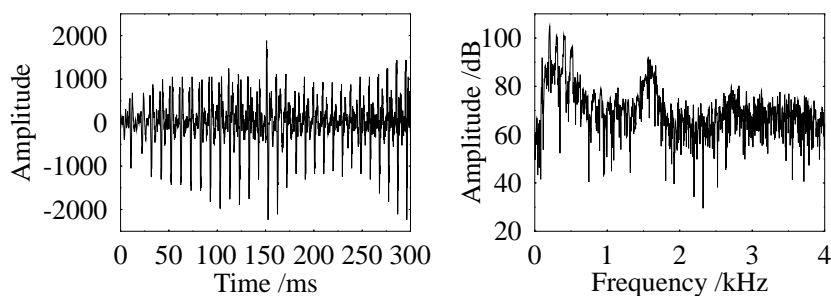


Figure 18.1: Example of a sustained voiced fricative /z/ present in ‘zoo’. Observing the frequency domain, the phoneme is clearly voiced beneath 1kHz and much more noisy above 1KHz.

being represented by a mixture of voiced and unvoiced excitation sources.

This chapter commences with Section 18.2 giving an overview of a MMBE coder. Section 18.3 details the filters which construct the multiband structure, and discusses the additional complexity they introduce. An augmented exposure of a MMBE encoder is given in Section 18.4, with a closer view of a MMBE decoder detailed in Section 18.5. Finally Section 18.6 presents and examines the addition of the MMBE to the LPC vocoder of Chapter 15 and the PWI-ZFE scheme described in Chapter 17.

18.2 Overview of Mixed-Multiband Excitation

The control structure of a MMBE model is shown in Figures 18.3 and 18.4 which are considered next. The corresponding steps can also be followed with reference to the encoder and decoder schematics shown in Figure 18.5. After LPC analysis has been performed on the 20ms speech frame pitch detection occurs in order to locate any evidence of voicing. A frame deemed unvoiced has the RMS of its LPC residual quantized and sent to the decoder.

Speech frames labelled as voiced are split into M frequency bands, with M constrained to be a constant value. These frequency bands generally have a bandwidth which contains an integer number of pitch related spectral needles, where in the ideal situation each frequency band would have a width of one pitch related spectral needle. However, in practical terms, due to coding efficiency constraints, each frequency band contains several pitch related needles. The lower the fundamental frequency, the higher the number of pitch related needles per frequency band. A consequence of the time-variant pitch period is the need for the time-variant adaptive filterbank, which generates the frequency bands, to be reconstructed every frame in both the encoder and decoder, as shown in Figure 18.5, thus increasing the computational costs. Every frequency band is examined for voicing, before being assigned a voicing strength which is quantized and sent to the decoder. Reproduction of the speech at the decoder requires knowledge of the pitch period, in order to reconstruct the

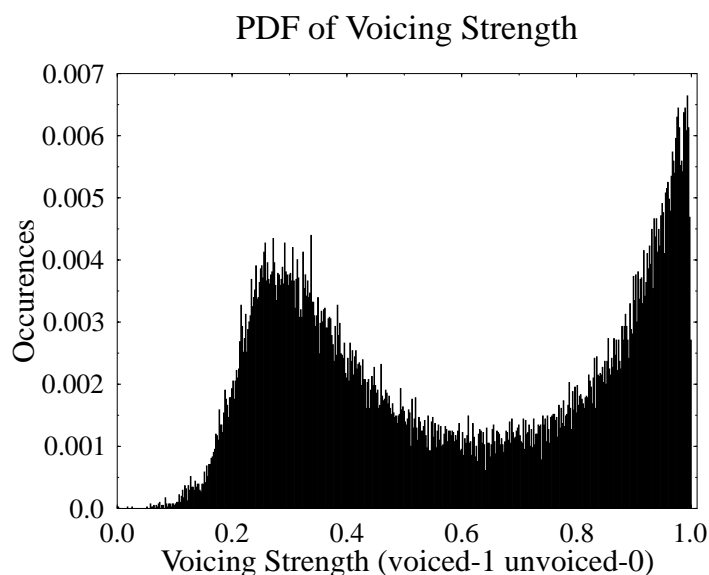


Figure 18.2: The distribution of voicing strengths for the training speech database of Table 14.1.

filterbanks of Figure 18.5(b), together with the voicing strength in each band. The voiced excitation must also be determined and its parameters have to be sent to the decoder.

At the decoder, following Figure 18.5(b), both unvoiced and voiced speech frames have a pair of filterbanks created. However, for unvoiced frames the filterbank is declared fully unvoiced with no pulses employed. For the voiced speech frames both voiced and unvoiced excitation sources are created.

Following Figure 18.4, both the voiced and unvoiced filterbanks are created using the knowledge of the pitch period and the number of frequency bands, M . For the voiced filterbanks the filter coefficients are scaled by the quantized voicing strengths determined at the encoder. A value of 1 represents full voicing, while a value 0 signifies a frequency band of noise, with values between these extremes representing a mixed excitation source. For the unvoiced filterbank the voicing strengths are adjusted, ensuring that the voicing strengths of each voiced and unvoiced frequency band combine to unity. This constraint maintains a combined resultant from the filterbanks that is spectrally flat over the entire frequency range. The mixed excitation speech is then synthesized, as shown in Figure 18.5(b), where the LPC filter determines the spectral envelope of the speech signal. The construction of the filterbanks is described in detail in Section 18.3.

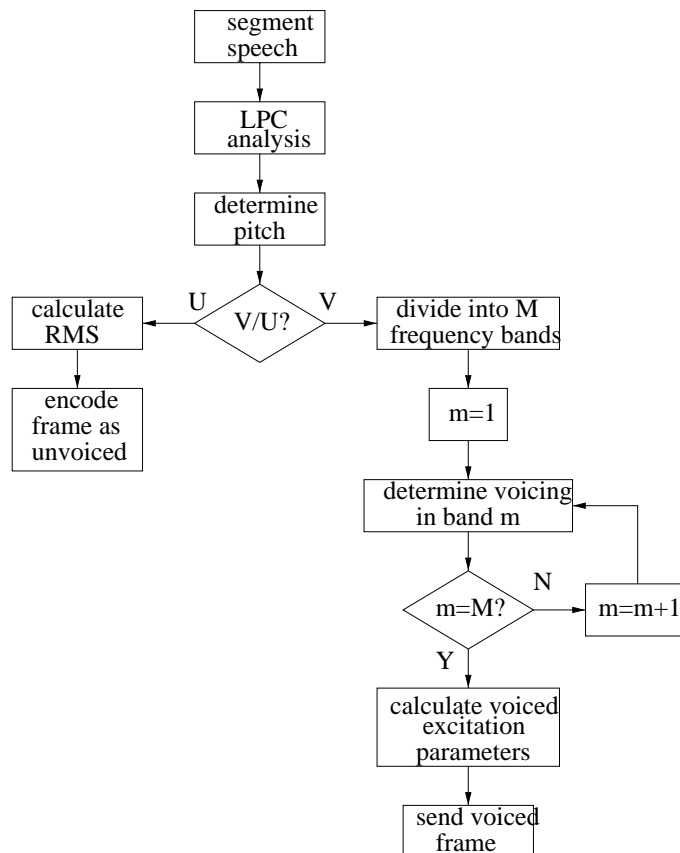


Figure 18.3: Control structure for a MMBE encoder.

18.3 Finite Impulse Response Filter

The success of MMBE is dependent on creating a suitable bank of filters. The filterbank should be capable of producing either fully voiced or unvoiced speech together with mixed speech. Two well-established techniques for producing filterbanks are Finite Impulse Response (FIR) filters and Quadrature Mirror Filters (QMFs), a type of FIR filter.

QMFs [354] are designed to divide a frequency spectrum in half, thus a cascade of QMFs can be implemented until the spectrum is divided into appropriate frequency bands. If a signal has a sampling frequency f_s , then a pair of QMFs will divide the signal into a band from 0 to $f_s/4$ and a band from $f_s/4$ to $f_s/2$. Both filters will have their 3dB point at $f_s/4$. The filterbank of our MMBE coder was not constructed from QMFs, since the uniform division of the frequency spectrum imposes restrictions on the shape of the filterbank.

FIR filters contain only a finite number of non-zero impulse response taps, thus, for

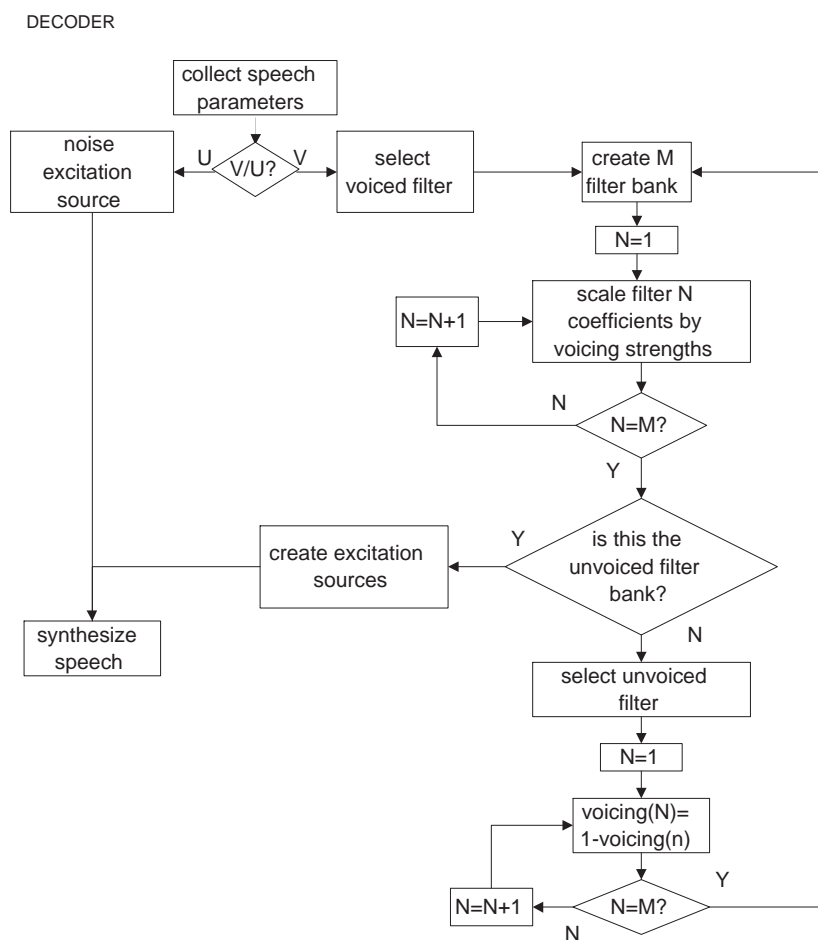
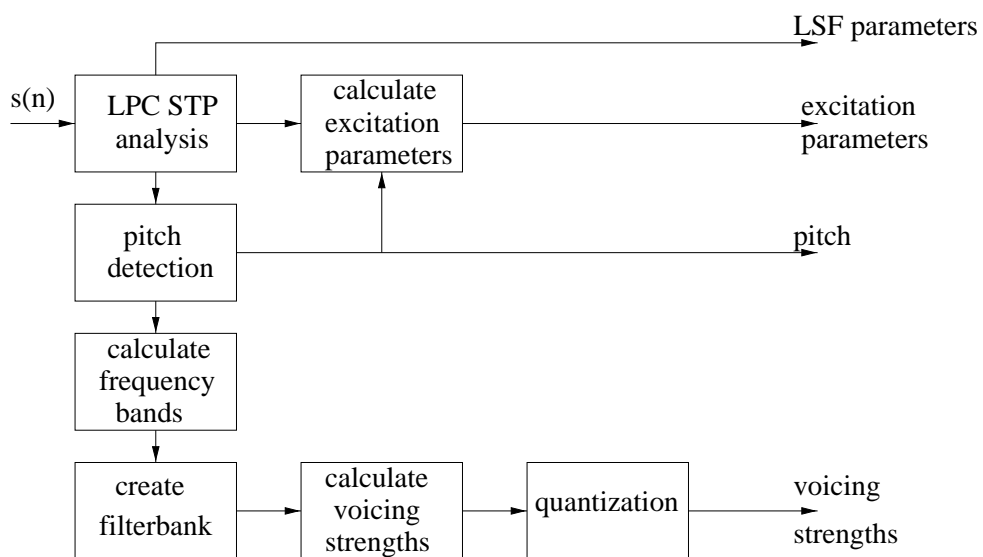
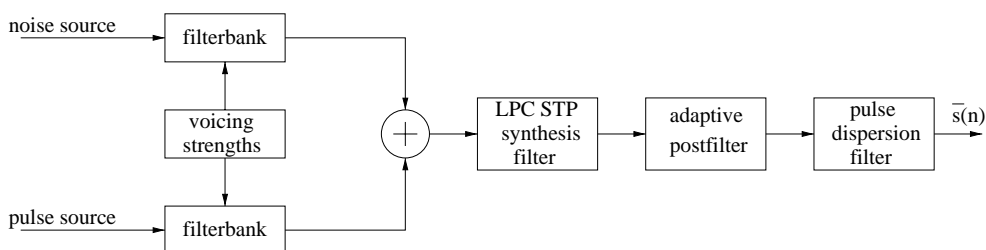


Figure 18.4: Control structure for a MMBE decoder.



(a) Encoder



(b) Decoder

Figure 18.5: Schematic of the a) encoder and b) decoder for a MMBE scheme.

18.3. FINITE IMPULSE RESPONSE FILTER**747**

a FIR filter of length K the impulse response is given by:

$$h_T(n) = \begin{cases} b_n & 0 \leq n \leq K-1 \\ 0 & \text{elsewhere} \end{cases} \quad (18.1)$$

where $h_T(n)$ is the impulse response of the filter and b_n are the filter coefficients. Using discrete convolution the filter's output signal is given by:

$$y_T(n) = \sum_{m=0}^{K-1} h_T(m) \cdot x_T(n-m) \quad (18.2)$$

where y_T is the filter output and x_T is the filter input. Computing the Z-transform of Equation 18.2, we arrive at the following filter transfer function:

$$H(z) = \sum_{m=0}^{K-1} h_T(m) z^{-m} \quad (18.3)$$

The impulse response of an ideal low pass filter transfer function $H(z)$ is the well known infinite duration sinc function given below:

$$h_T(n) = \frac{1}{\pi n r_c} \sin(2\pi n r_c) \quad (18.4)$$

where r_c is the cutoff frequency which has been normalized to $f_s/2$. In order to create a windowed ideal FIR low pass filter we invoke a windowing function $w(n)$, which is harnessed as follows:

$$h_T(n) = \frac{1}{\pi n r_c} w_{ham}(n) \sin(2\pi n r_c) \quad (18.5)$$

where $w_{ham}(n)$ was chosen in our implementation to be the Hamming window given by:

$$w_{ham}(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{K}\right) \quad (18.6)$$

with K being the filter length. In order to transform the low pass filter to a bandpass filter, h_T^{BP} , the ideal windowed low pass filter, h_T^{LP} is scaled by the expression [451]:

$$h_T^{BP}(n) = h_T^{LP}(n) \cos\left(2\pi n \left(\frac{r_l + r_u}{2}\right)\right) \quad (18.7)$$

where r_l is the lower normalized bandpass frequency and r_u is the upper normalized bandpass frequency.

A filterbank consists of the low pass filter together with the bandpass filters such that the entire frequency range is covered. Thus, as demonstrated in Figure 18.6, the filterbank contains both a low pass filter and bandpass filters in its constitution.

Following this overview of MMBE, the extra processes required by MMBE within a speech encoder are discussed in the next Section.

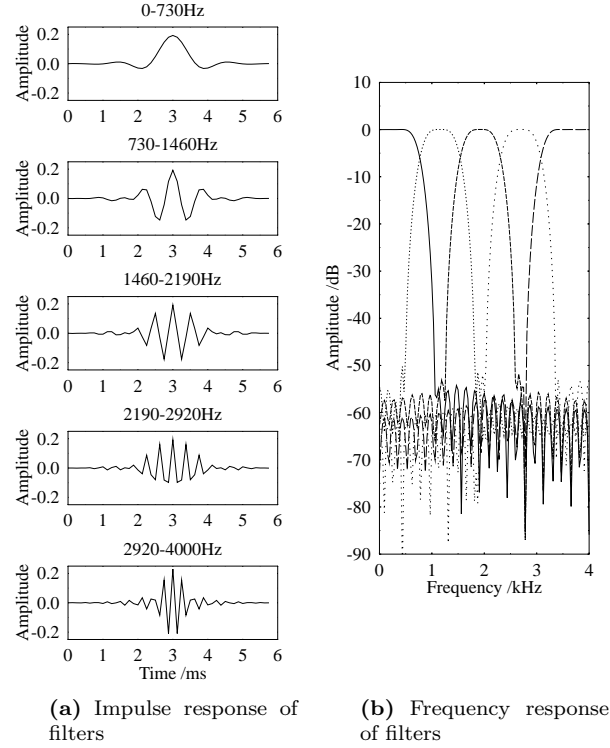


Figure 18.6: The a) impulse responses and b) frequency responses for a filterbank constructed from a lowpass and four bandpass filters. They have frequency ranges $0 \rightarrow 730\text{Hz}$, $730 \rightarrow 1460\text{Hz}$, $1460 \rightarrow 2190\text{Hz}$, $2190 \rightarrow 2920\text{Hz}$ and $2920 \rightarrow 4000\text{Hz}$. A filter order of 47 was used.

18.4 Mixed-Multiband Excitation Encoder

At the encoder the task of the filterbank is to split the frequency band and facilitate the determination of the voicing strengths in each frequency band. In order to accommodate an integer number of the spectral domain pitch related needles, each frequency band's bandwidth is a multiple of the fundamental frequency. The total speech bandwidth, $f_s/2$, is occupied by a $N_n \cdot F0 \cdot M$ number of pitch related needles, where f_s is the sampling frequency, $F0$ is the fundamental frequency and M is the number of bands in the filterbank, while N_n is the number of needles for each subband, which can be expressed as [452]:

$$N_n = \frac{f_s/2}{M \cdot F0} \quad (18.8)$$

The resultant N_n value is rounded down to the nearest integer. Any remaining

frequency band between $f_s/2$ and the final filter cutoff frequency is assumed unvoiced.

For example, with a sampling frequency of 8kHz and a filterbank design having five bands the number of harmonics in each band can be determined. For a fundamental frequency of 100Hz it follows that:

$$N_n = \frac{4000}{100 \times 5} = 8 \quad (18.9)$$

implying that there will be eight pitch needles for each subband. Similarly, for a fundamental frequency of 150Hz, we have:

$$N_n = \frac{4000}{150 \times 5} = 5.33 \quad (18.10)$$

Thus each band will contain five pitch needles, with the frequencies 3750 to 4000Hz being incorporated in the upper frequency band.

The method of dividing the frequency spectrum as suggested by Equation 18.8 is not a unique solution. It would be equally possible to increase the bandwidth of the higher filters due to the human ear's placing less perceptual emphasis on these regions. However, the above pitch dependent, but even spread of the frequency bands allows a simple division of the frequency spectrum. Since the decoder reconstructs the filter from $F0$ no extra side information requires transmission.

18.4.1 Voicing Strengths

For every voiced speech frame the input speech is passed through each filter in the filterbank, in order to locate any evidence of voicing in each band. Figure 18.7 shows the transfer function of the filterbank created and the filtered speech in both the time and frequency domain. Observing the top of Figure 18.7(a), below 3kHz the original spectrum appears predominantly voiced, whereas above 3kHz it appears more unvoiced, as shown by the periodic and aperiodic spectral fine structure present. The corresponding time domain signal waveforms, of Figure 18.7(b) seem to contain substantially attenuated harmonics of the fundamental frequency $F0$, although the highest two frequency bands appear more noise-like.

The voicing strength is found in our coder using several methods [400], since if the voicing is inaccurately calculated the reconstructed speech will contain an excessive 'buzz' or 'hiss', that is too much periodicity or excessive noise, respectively. Initially the voicing strength, v_s , is found using the normalized pitch-spaced filtered waveform correlation [400]:

$$v_s = \frac{\sum_{n=0}^{FL-1} f(n) * f(n-P)}{\sqrt{\sum_{n=0}^{FL-1} f(n)^2 \sum_{n=0}^{FL-1} f(n-P)^2}} \quad (18.11)$$

where $f(n)$ is the filtered speech of a certain bandwidth, FL is the frame length and P is the pitch period for the speech frame. However, at the higher frequencies the correlation can be very low even for voiced speech. The time domain envelope of the filtered speech will be a better indication of voicing [400], as demonstrated by Figure 18.8.

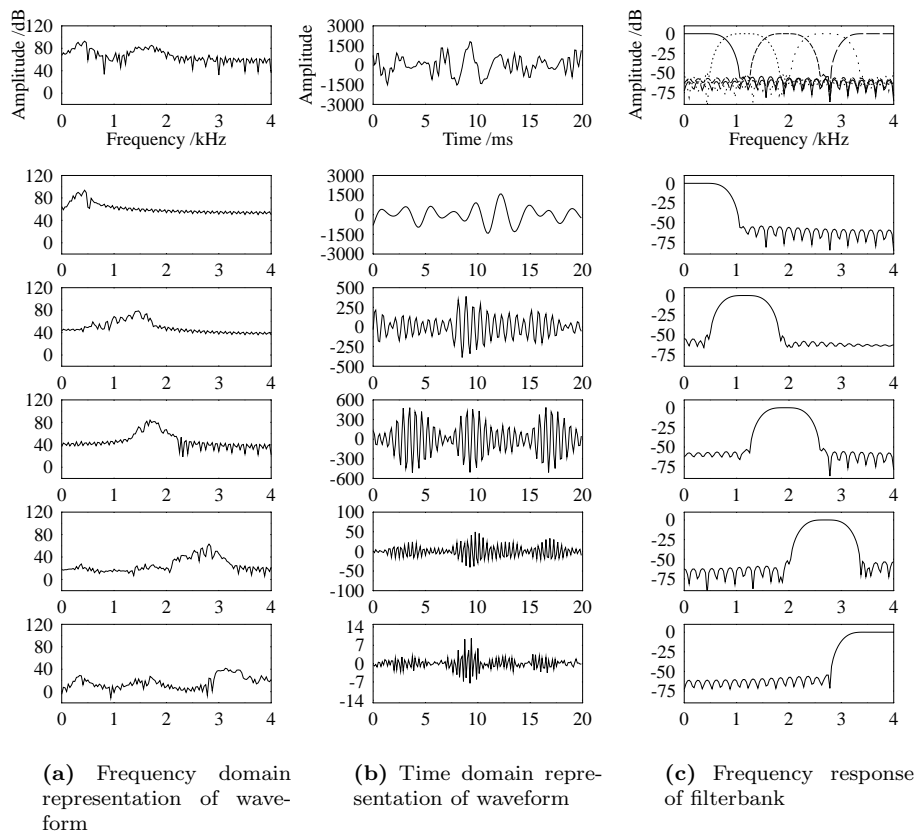


Figure 18.7: The a) frequency and b) time domain representation of the original waveform AM1 when uttering diphthong /aɪ/ in ‘wires’ together with the filtered waveform. c) The frequency responses of the filterbank are also shown. A filter order of 47 was used.

The envelope of the bandpass filtered speech is found through low pass filtering the full-wave rectified filtered speech signal. The one-pole low pass filtered rectified bandpass signal is given by:

$$f(n) = \frac{1}{1 + 2\pi f_c / f_s} \cdot [2\pi \frac{f_c}{f_s} s(n) + f(n-1)] \quad (18.12)$$

where f_c is the cutoff frequency, $s(n)$ is the input signal of the filter, $f(n)$ is the output signal of the filter and f_s is the sampling frequency. The cutoff frequency was taken to be 500Hz, since this is just above the highest expected fundamental frequency. The voicing strength, v_s , is then calculated using Equation 18.11 for the low pass filtered, rectified bandpass signal.

Subsequently, each frequency band is assigned the largest calculated voicing strength

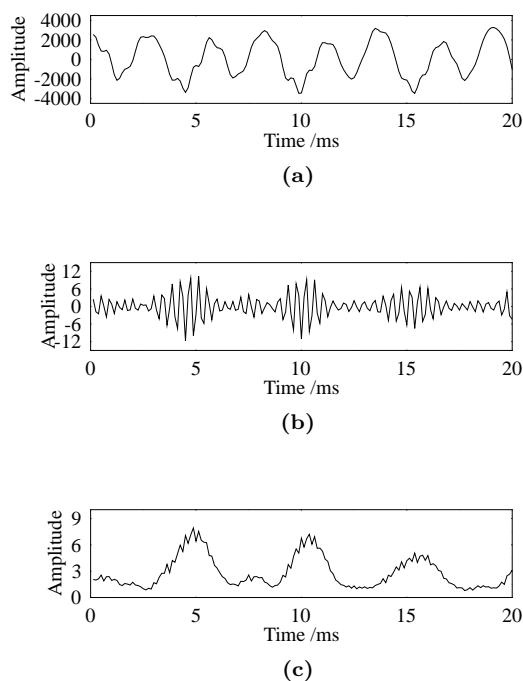


Figure 18.8: Time domain waveforms of a) the original speech, b) the bandpass filtered speech and c) the envelope of the bandpass filtered speech.

achieved from the original bandpass signal or the low pass filtered rectified bandpass signal. The PDF of the selected voicing strengths for a 20-band filterbank is given in Figure 18.9 for the training database. The graph represents all the voicing strengths recorded in every frequency band, providing sufficient fine resolution training data for the Max-Lloyd quantizer to be used.

The PDF for the voicing strength values was passed to the Max-Lloyd quantizer described in Section 15.4. The Max-Lloyd quantizer allocates eight levels for the voicing strengths using a total of 3-bits, with level 0 constrained to be 0.2 and level 8 constrained to be 1. If level 0 was assigned to be 0 the quantizer would be too biased towards the lower valued voicing strengths. The same quantizer is used to encode every frequency band, producing the SNR values for a 1-band to 15-band MMBE scheme given in Figure 18.10, where the speech files AM1, AM2, AF1, AF2, BM1, BM2, BF1 and BF2 were used to test the quality of the MMBE quantizer.

This Section has detailed a range of processes invoked in a speech encoder due to MMBE, while in the next Section procedures required by the MMBE decoder are revealed.

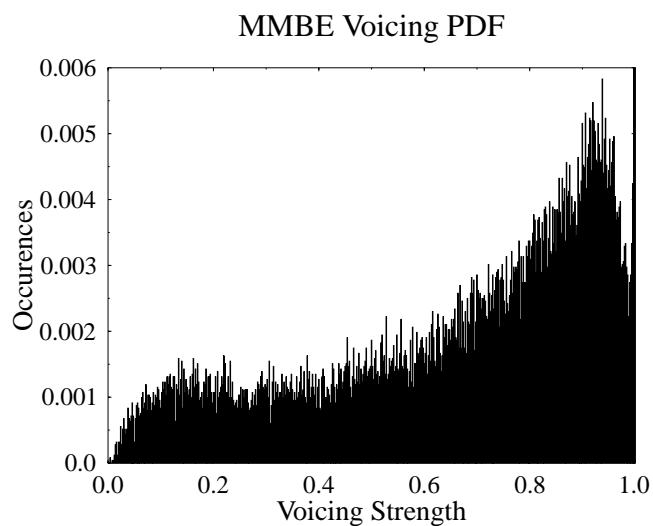


Figure 18.9: The PDF of the voicing strengths for an 20-band filterbank using the database of Table 14.1.

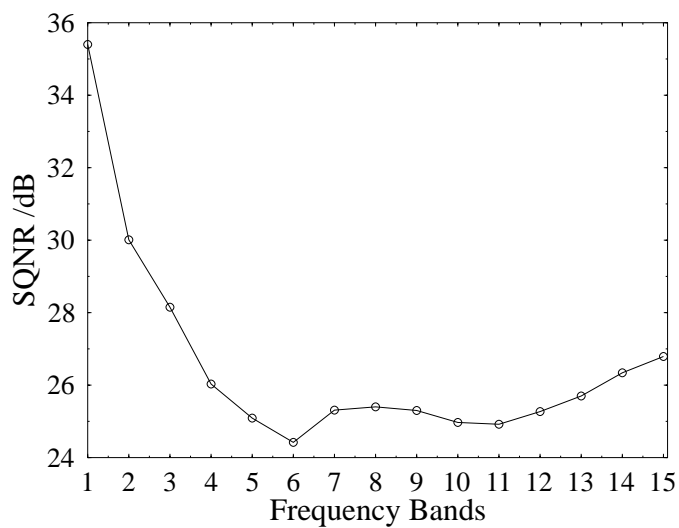


Figure 18.10: SNR values, related to the quantized and unquantized voicing strengths, achieved after the voicing levels in the respective frequency bands are 3-bit quantized for the MMBE coder.

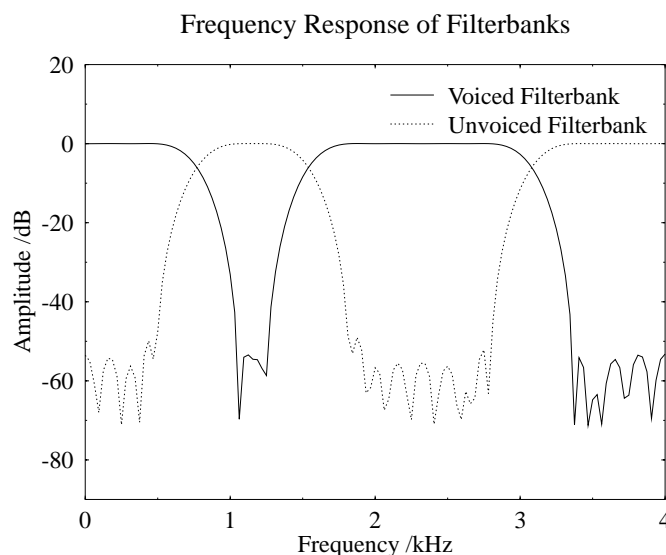


Figure 18.11: Constructed voiced and unvoiced filterbanks for the MMBE decoder. Displayed for a 5-band model, with three voiced and two unvoiced bands. Using a filter of order 47.

18.5 Mixed-Multiband Excitation Decoder

In the MMBE scheme at the decoder of Figure 18.4 and 18.5(b), for voiced speech two versions of the filterbank are constructed, which will be justified below. Subsequently both voiced and unvoiced excitation are passed through these filterbanks and onto the LPC synthesis filter, in order to reproduce the speech waveform.

Explicitly, the power of the filterbank generating the voiced excitation is scaled by the quantized voicing strength, while the filterbank producing the unvoiced excitation is scaled by the difference between unity and the voicing strength. This is performed for each of the frequency bands of the filterbank. Once combined the resultant filterbanks produce an all-pass filter over the 0 to 4000Hz frequency range, as demonstrated in Figure 18.11. The filterbanks are designed to allow complete voicing, pure noise, or any mixture of the voiced and unvoiced excitation. As specified in Section 18.4 any frequency in the immediate vicinity of 4kHz which was not designated a voicing strength is included in the upper most frequency band. From the knowledge of the fundamental frequency F_0 and the number of bands M the decoder computes N_n , the number of pitch-related needles in each frequency band. Thus, with the normalized cut-off frequencies known the corresponding impulse response can be inferred from Equations 18.5 and 18.7.

For both voiced and unvoiced speech frames the $(1 - v_s)$ scaled noise excitation is

Parameter	Values			
	2-band MMBE	5-band MMBE	3-band MMBE PWI-ZFE	13-band MMBE PWI-ZFE
α_{pf}	0.75	0.80	0.85	0.85
β_{pf}	0.45	0.55	0.55	0.50
μ_{pf}	0.60	0.50	0.60	0.60
γ_{pf}	0.50	0.50	0.50	0.50
g_{pf}	0.00	0.00	0.00	0.00
ξ_{pf}	0.99	0.99	0.99	0.99

Table 18.1: Appropriate adaptive postfilter values for the MMBE speech coders examined in Section 18.6.

passed to the unvoiced filterbank. The voiced excitation is implemented with either pulses from the LPC vocoder, as detailed in Section 15.5, or using the PWI-ZFE function detailed in Section 17.5. Then after scaling by v_s the excitation is passed to the voiced filterbank. The filtered signals are combined and passed to the LPC STP filter for synthesis.

In Figure 18.12 the process of selecting the portion of the frequency spectrum that is voiced and unvoiced is shown. Figure 18.12(a) shows the original speech spectrum with its LPC STP residual signal portrayed in Figure 18.12(b). Figure 18.12(c) and Figure 18.12(d) represent the voiced and unvoiced excitation spectra, respectively. From Figure 18.12(f) it can be seen that beneath 2kHz the classification is voiced, while above 2kHz it has been classified unvoiced. Lastly, Figure 18.12(e) demonstrates the synthesized frequency spectrum.

18.5.1 Adaptive Postfilter

The adaptive postfilter from Section 15.6 was used for the MMBE speech coders, with Table 18.1 detailing the optimized parameters for each MMBE speech coder detailed in the next Section. Following adaptive postfiltering the speech is passed through the pulse dispersion filter of Figure 15.19. In the next Section we now consider the issues of algorithmic complexity.

18.5.2 Computational Complexity

The additional computational complexity introduced by a MMBE scheme in both the encoder and decoder is given in Table 18.2 and Figure 18.13. From Table 18.2 it can be seen that at the encoder the complexity is dominated by the process of filtering the speech into different bands, while at the decoder the MMBE filtering process is dominant. In Figure 18.13 frequency band schemes between 1-band and 15-band are considered.

Following this description of the MMBE process, the reconstructed speech is examined when MMBE is added to both the benchmark LPC vocoder of Chapter 15 and the PWI-ZFE coder of Chapter 17.

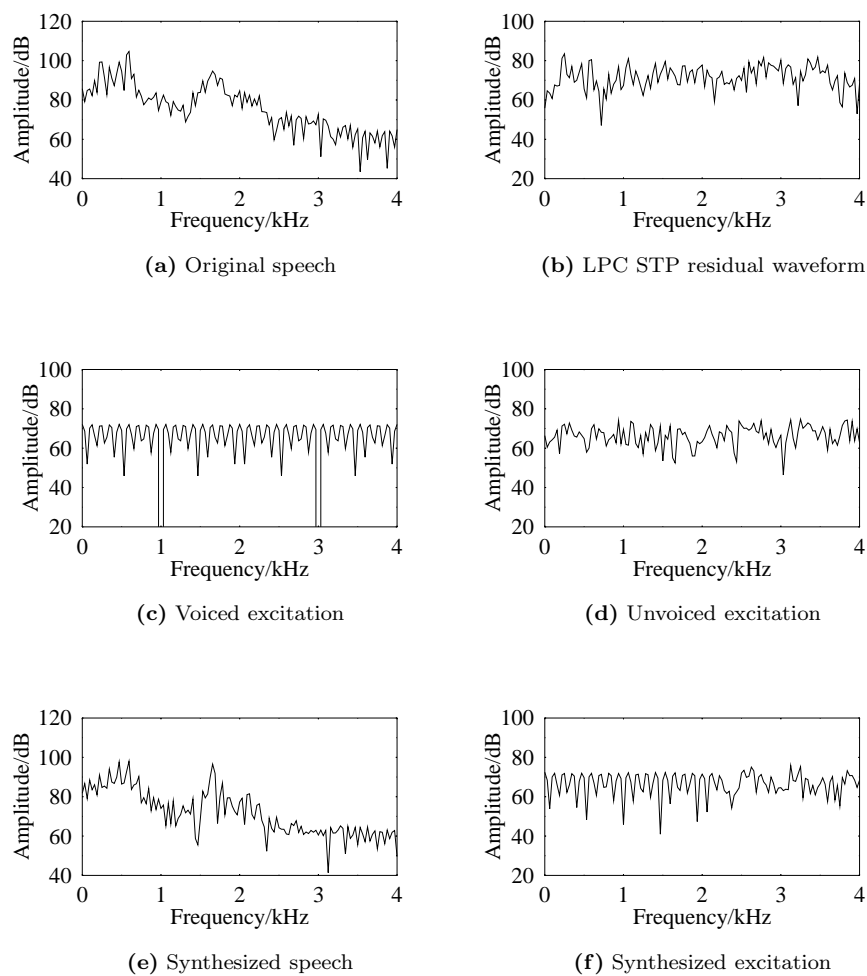


Figure 18.12: An example of the MMBE process for a 20ms speech frame from the testfile AM1 when uttering the back vowel /u/ in 'should'. The a) original and e) synthesized frequency spectrum is demonstrated, along with the b) original and f) synthesized excitation spectra, also shown are the c) voiced and d) unvoiced excitation spectra.

Procedure	2-band /MFLOPS	5-band /MFLOPS
Encoder		
Create filterbank	0.02	0.05
Filter speech into bands	1.54	3.07
Find voicing strengths	0.35	0.88
Decoder		
Create filterbank	0.02	0.05
Filter excitation sources	3.11	7.77

Table 18.2: Additional computational complexity introduced at the encoder and decoder by the MMBE scheme, for a 2-band and 5-band arrangement.

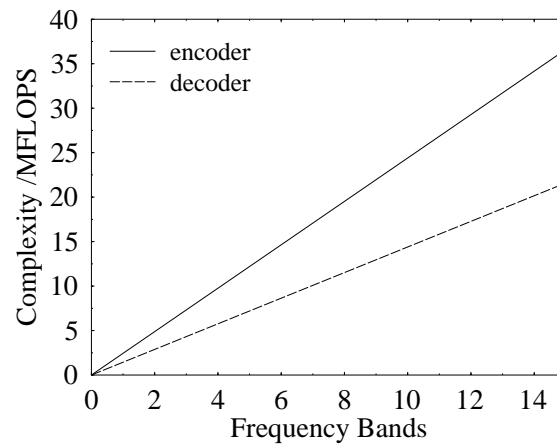


Figure 18.13: The computational complexity in the MMBE encoder and decoder for different numbers of frequency bands.

18.6 Performance of the Mixed-Multiband Excitation Coder

This Section discusses the performance of the benchmark LPC vocoder, in Chapter 15, and the PWI-ZFE coder in Chapter 17, with the addition of MMBE. Both a 2-band and a 5-band MMBE were added to the LPC vocoder, as detailed in Section 18.6.1, creating speech coders operating at 1.85kbps and 2.3kbps, respectively. For the PWI-ZFE coder only a 3-band MMBE was added, as detailed in Section 18.6.2, producing a 2.35kbps speech coder.

18.6. PERFORMANCE OF THE MIXED-MULTIBAND EXCITATION CODER⁷⁵⁷

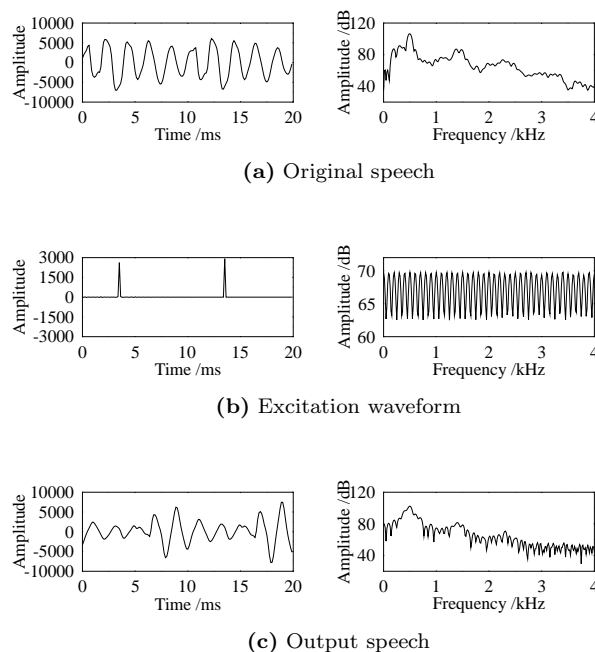


Figure 18.14: Time and frequency domain comparison of the a) original speech, b) **2-band MMBE** waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the mid vowel /ɜ:/ in the utterance ‘**work**’ for the testfile **BM1**. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

18.6.1 Performance of a Mixed-Multiband Excitation Linear Predictive Coder

The MMBE scheme, as detailed in this chapter was added to the basic LPC vocoder described in Chapter 15, with the speech database described in Table 14.1 used to assess the coder’s performance. The time- and frequency-domain plots for individual 20ms frames of speech are given in Figures 18.14, 18.15 and 18.16 for a 2-band MMBE model, while Figures 18.17, 18.18 and 18.19 display the corresponding results for a 5-band MMBE model. Both Figures 18.14 and 18.17 represent the same speech segment as Figures 15.21 and Figures 17.13, while Figures 18.15 and 18.18 represent the same speech segment as Figure 15.22 and Figure 17.14, and Figures 18.16 and 18.19 represent the same speech segment as Figure 15.23 and Figure 17.15. Initially, the performance of a 2-band MMBE scheme is studied.

Figure 18.14 displays the performance of a 20ms speech frame from the testfile **BM1**. For this speech frame Figure 18.14(b) shows that the entire frequency spectrum is considered voiced, thus the reproduced speech waveform is identical to Figure 15.21.

Figure 18.15 is an utterance from the testfile **BF2**, where observing Figure 18.15(b)

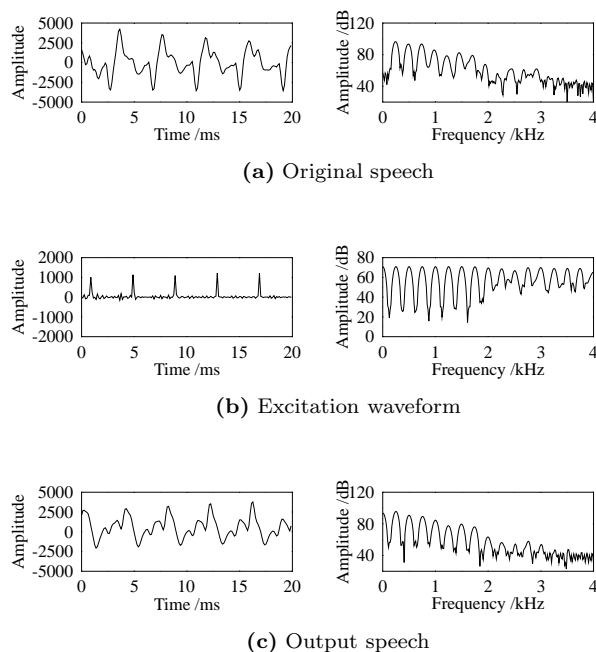


Figure 18.15: Time and frequency domain comparison of the a) original speech, b) **2-band MMBE** waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the liquid /r/ in the utterance ‘**rice**’ for the testfile **BF2**. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

above 2kHz a mixture of voiced and unvoiced excitation is harnessed. From Figure 18.15(c) it can be seen that the presence of noise above 2kHz produces a better representation of the frequency spectrum than Figure 15.22(c).

Figure 18.16 is a 20ms speech frame from the testfile BM2 for the nasal /n/ in the utterance ‘thrown’. Similarly to Figure 18.15, the frequency spectrum above 2kHz is modelled by purely unvoiced excitation. Figures 18.15 and 18.16 demonstrate that many speech waveforms contain both voiced and unvoiced components, thus, they emphasize the need for a speech coder which can incorporate mixed excitation.

Through informal listening a comparison of the synthesized speech from an LPC vocoder with and without MMBE can be made. The introduction of the MMBE removes a significant amount of the ‘buzz’ inherent in LPC vocoder models, producing more natural sounding speech. Occasionally a background ‘hiss’ is introduced into the synthesized speech, which is due to the coarse resolution of the frequency bands in a 2-band MMBE scheme. Additionally, pairwise-comparison tests, detailed in Section 20.2, were conducted to compare the speech quality from the 1.9kbps PWI-ZFE speech coder of Chapter 17 with the 2-band MMBE LPC scheme. These pairwise-comparison tests showed that 30.77% of listeners preferred the PWI-ZFE

18.6. PERFORMANCE OF THE MIXED-MULTIBAND EXCITATION CODER 759

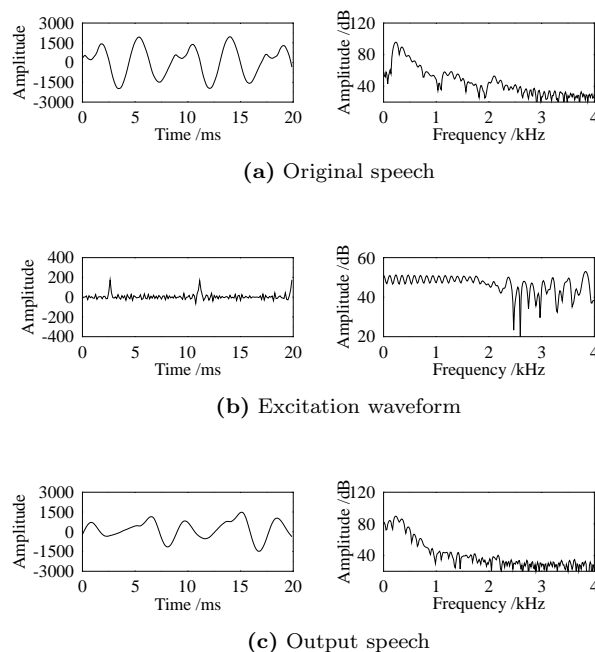


Figure 18.16: Time and frequency domain comparison of the a) original speech, b) **2-band MMBE** waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the nasal /n/ in the utterance ‘**thrown**’ for the testfile **BM2**. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

speech coder, with 23.07% of listeners preferring the 2-band MMBE LPC scheme and 46.16% having no preference.

A 5-band MMBE scheme was also implemented in the context of the LPC vocoder, which with an increased number of voicing decisions should produce better quality synthesized speech than the 2-band MMBE model.

For Figure 18.14 the addition of the extra three extra frequency bands is shown in Figure 18.17 for a speech frame in the testfile BM1. From Figure 18.17(b) it can be seen that the extra frequency bands produce a mixture of voiced and unvoiced speech above 3kHz, where for the 2-band MMBE model the entire frequency spectrum was fully voiced.

Figure 18.18 portrays the speech frame shown in Figure 18.15 from the BF2 testfile, but with an extra three frequency bands. For this speech frame the additional three frequency bands have no visible effect.

Figure 18.19 displays a speech frame from the testfile BM2 with 5-band MMBE and can be compared with Figure 18.16. For this speech frame the addition of three frequency bands produces fully unvoiced speech above 800Hz, as shown in Figure 18.19(b), with the effect on the synthesized speech visible in the frequency domain

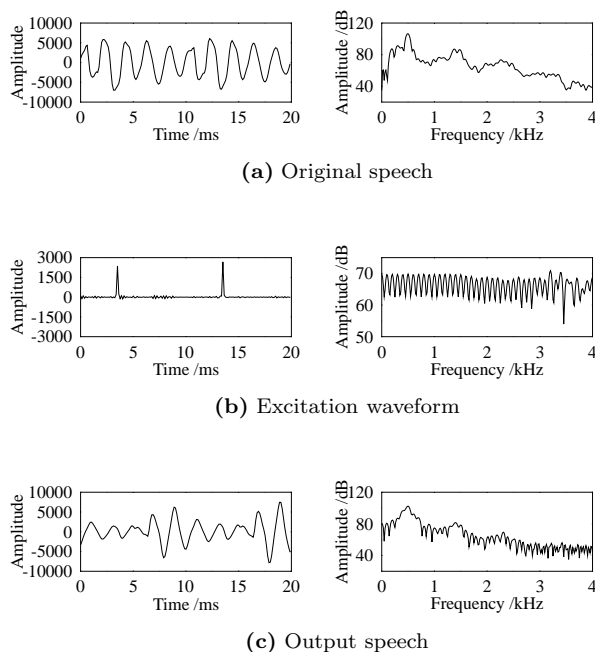


Figure 18.17: Time and frequency domain comparison of the a) original speech, b) 5-band MMBE waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the mid vowel /ɜ:/ in the utterance ‘*work*’ for the testfile BM1. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

of Figure 18.19(c).

With informal listening tests it was found that the addition of an extra three decision bands to the MMBE scheme has little perceptual effect. It is possible that inherent distortions caused by the LPC vocoder model are masking the improvements. The bit allocation for an LPC vocoder with either a 2-band or 5-band MMBE scheme is given in Table 18.3. The voicing strength of each decision band is quantized with a 3-bit quantizer as described in Section 18.4, thus, adding 0.15kbps to the overall bit rate of the coder. The computational complexity of the LPC speech vocoder with 2- and 5-band MMBE is given in Table 18.4, where the complexity is dominated by the MMBE function.

In the next Section a 3-band MMBE scheme is incorporated into the PWI-ZFE coder of Chapter 17.

18.6. PERFORMANCE OF THE MIXED-MULTIBAND EXCITATION CODER761

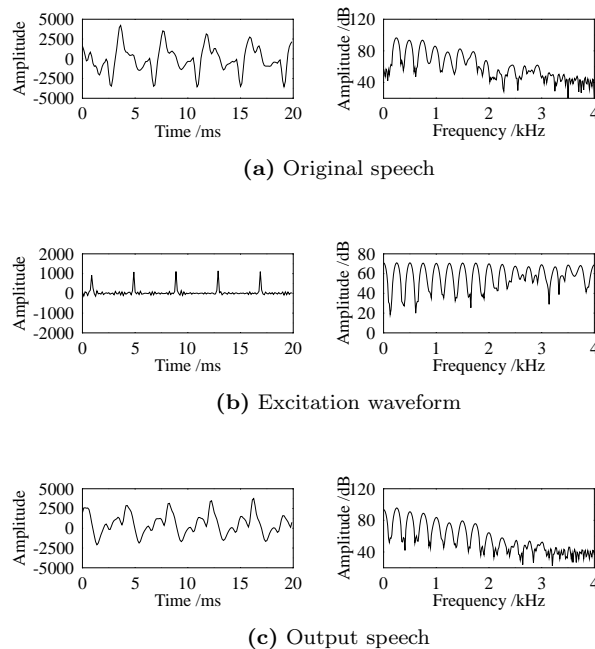


Figure 18.18: Time and frequency domain comparison of the a) original speech, b) **5-band MMBE** waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the liquid /r/ in the utterance ‘**rice**’ for the testfile **BF2**. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

parameter	2-band	5-band
LSFs	18	18
V/U flag	1	1
RMS value	5	5
Pitch	7	7
Voicing strengths	2×3	5×3
total/20ms	37	46
bit rate	1.85kbps	2.30kbps

Table 18.3: Bit allocation table for the LPC vocoder voiced frames with 2-band and 5-band MMBE.

Operation	2-band complexity MFLOPS	5-band complexity MFLOPS
Pitch detector	2.67	2.67
MMBE filtering	1.91	4.00
Total	4.58	6.67

Table 18.4: Total computational complexity for a basic LPC vocoder encoder with either a 2-band or 5-band MMBE model.

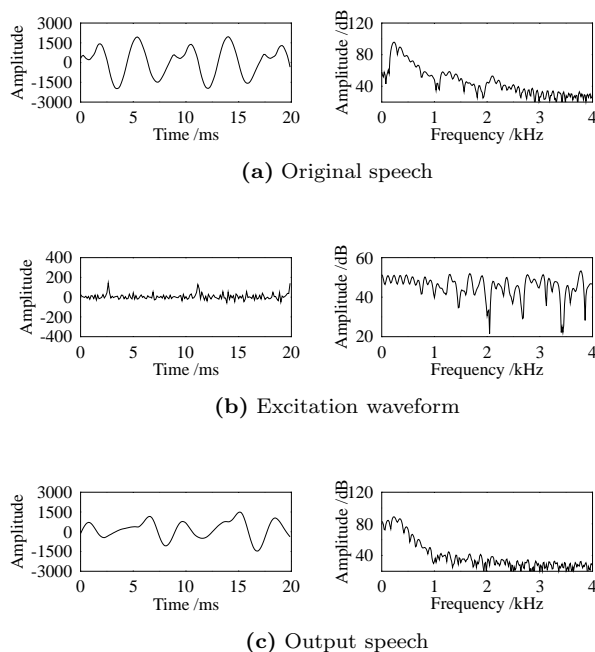


Figure 18.19: Time and frequency domain comparison of the a) original speech, b) 5-band MMBE waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the nasal /n/ in the utterance ‘**thrown**’ for the testfile BM2. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

18.6.2 Performance of a Mixed-Multiband Excitation and Zinc Function Prototype Excitation Coder

The MMBE scheme, detailed in this chapter was also added to the PWI-ZFE coder described in Chapter 17. Again, the speech database described in Table 14.1 was used to assess the coder’s performance. The time and frequency domain plots for individual 20ms frames of speech are given in Figures 18.20, 18.21 and 18.22 for a 3-band MMBE excitation model. These are the speech frames consistently used to consider the performance of the coders, thus, can be compared with Figures 17.13, 17.14 and 17.15, respectively, together with those detailed in Table 20.2.

Figure 18.20 displays the performance of a 3-band MMBE scheme incorporated in the PWI-ZFE speech coder for a speech frame from the testfile BM1. Observing the frequency domain of Figure 18.20(b), a small amount of unvoiced speech is present above 2.5kHz. The changes this noise makes to the synthesized speech is visible in the frequency domain of Figure 18.20(c).

Similarly to Figure 18.20, for the speaker BF2 Figure 18.21 displays evidence of noise above 2.5kHz. This noise is again visible in the frequency domain of Figure 18.21(c).

18.6. PERFORMANCE OF THE MIXED-MULTIBAND EXCITATION CODER⁷⁶³

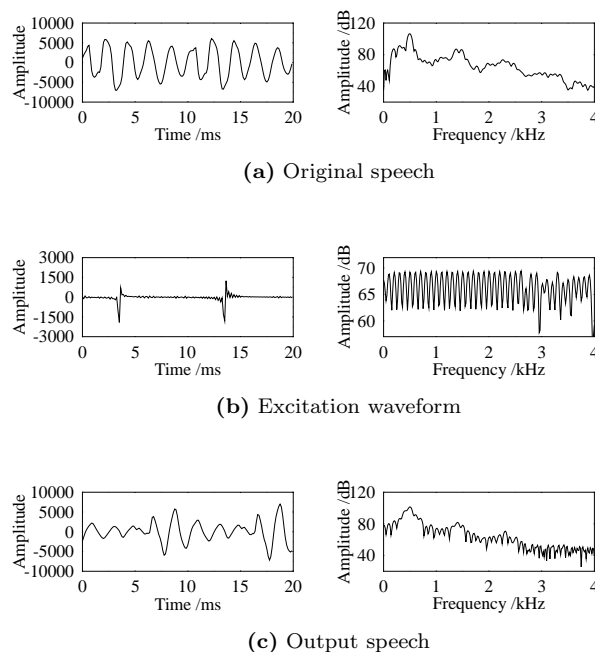


Figure 18.20: Time and frequency domain comparison of the a) original speech, b) **3-band MMBE ZFE** waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the mid vowel /ɜ:/ in the utterance ‘**work**’ for the testfile **BM1**. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

The introduction of a 3-band MMBE scheme to the PWI-ZFE speech coder has a more pronounced effect in the context of the testfile BM2, as shown in Figure 18.22. From Figure 18.22(b) it can be seen that above 1.3kHz the frequency spectrum is entirely noise. In the time domain much more noise is evident in the excitation waveform than for either Figure 18.20(b) or 18.21(b).

Through informal listening to the PWI-ZFE coder, any audible improvements achieved by the addition of 3-band MMBE can be assessed. The MMBE removes much of the ‘buzziness’ from the synthesized speech, which particularly improves the speech quality of the female speakers. Occasionally, the MMBE introduces ‘hoarseness’, indicative of too much noise, especially to the synthesized speech of male speakers, but overall the MMBE improves speech quality at a slightly increased bit rate and complexity. Pairwise-comparison tests, detailed in Section 20.2, were conducted between the 2.35kbps 3-band MMBE PWI-ZFE speech coder and the 2.3kbps 5-band MMBE LPC scheme. These pairwise-comparison tests showed that 64.10% of listeners preferred the 3-band MMBE PWI-ZFE speech coder, with 5.13% of listeners preferring the 5-band MMBE LPC scheme and 30.77% having no preference.

As stated previously, each decision band introduces an additional 0.15kbps to the

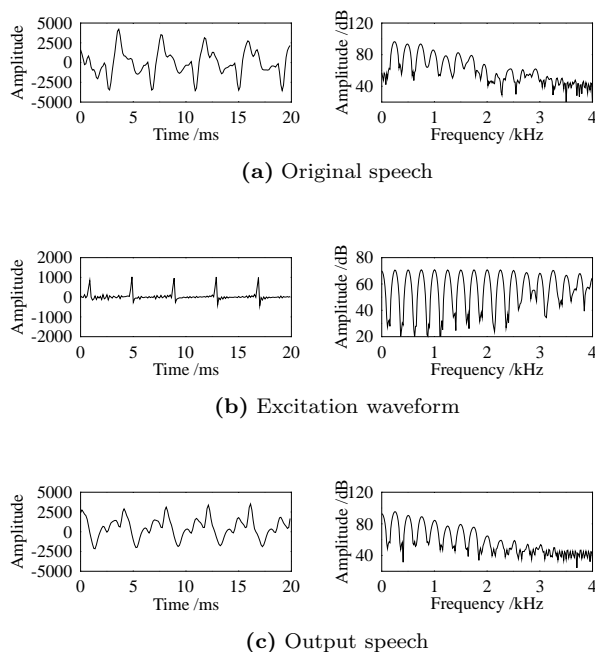


Figure 18.21: Time and frequency domain comparison of the a) original speech, b) **3-band MMBE ZFE** waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the liquid /r/ in the utterance ‘rice’ for the testfile **BF2**. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

overall bit rate of a speech coder. Hence, Table 18.5 shows that the addition of the MMBE scheme to the PWI-ZFE coder produced an overall bit rate of 2.35kbps.

The computational complexity of the PWI-ZFE speech vocoder with 3-band MMBE is given in Table 18.6, which is dominated by the filtering procedures involved in the MMBE process and the ZFE optimization process.

In this Section two schemes have been described which operate at similar bit rates, namely the LPC vocoder with 5-band MMBE operating at 2.3kbps, and the PWI-ZFE coder incorporating 3-band MMBE transmitting at 2.35kbps. With informal listening tests it was found that the PWI-ZFE coder with 3-band MMBE produced synthesized speech with slightly preferred perceptual qualities, although the quality of the reproduced speech was not dissimilar.

18.6. PERFORMANCE OF THE MIXED-MULTIBAND EXCITATION CODER765

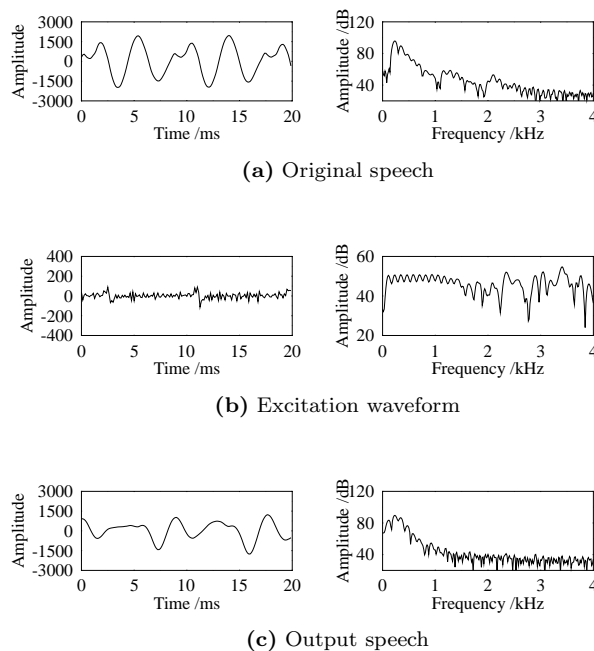


Figure 18.22: Time and frequency domain comparison of the a) original speech, b) **3-band MMBE ZFE** waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the nasal /n/ in the utterance ‘**thrown**’ for the testfile **BM2**. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

parameter	3-band	13-band
LSFs	18	18
v/u flag	1	1
Pitch	7	7
A_1	6	6
B_1	6	6
Voicing strengths	3×3	13×3
total/20ms	47	77
bit rate	2.35kbps	3.85kbps

Table 18.5: Bit allocation table for voiced frames in a 3-band and 13-bands MMBE PWI-ZFE speech coder.

Operation	3-band complexity /MFLOPS
Pitch detector	2.67
MMBE filtering	2.05
ZFE minimization	11.46
Total	16.18

Table 18.6: Total computational complexity for a PWI-ZFE coder with a 3-band MMBE arrangement.

18.7 A Higher Rate 3.85kbps Mixed-Multiband Excitation Scheme

In Sections 18.6.1 and 18.6.2 MMBE schemes operating at different bit rates have been investigated. The varying bits rates were achieved by either altering the excitation or by varying the number of frequency bands employed in the model. The nature of the pitch dependent filterbank, with the filterbank being reconstructed every frame, permits simple conversion between the number of frequency bands. Following the multiple ZFE investigation of Section 17.11 an MMBE scheme operating at 3.85kbps, incorporating a single ZFE, was implemented. The bit rate of 3.85kbps is close to the bit rate of the PWI-ZFE speech coder with three ZFEs of Chapter 17, allowing comparisons between the two techniques at a higher bit rate. The bit rate of 3.85kbps was achieved with the speech spectrum split into 13 bands, each scalar quantized with 3 bits as described in Section 18.4.

The performance for an MMBE-ZFE scheme at 3.85kbps is shown in Figure 18.23, Figure 18.24 and Figure 18.25, which can be compared with Figure 17.20, Figure 17.21 and Figure 17.22 showing the three pulse ZFE speech coder. Additional pertinent comparisons can be made with the Figures detailed in Table 20.2.

For a speech frame from the testfile BM1 displayed in Figure 18.23 the frequency spectrum is still predominantly voiced, with noise being added only above 2.7kHz. For this speech frame the MMBE extension to the PWI-ZFE model performs better than adding extra ZFE pulses, since as shown in Figure 17.20 these extra ZFE pulses introduced pitch doubling.

Figure 18.24 shows a frame of speech from the testfile BF2. For this speech frame Figure 18.24(b) shows that up to 1kHz the speech is voiced, between 1-2kHz a mixture of voiced and unvoiced speech is present in the spectrum, between 2-3kHz the speech is predominantly voiced, while above 3kHz only noise is present in the frequency spectrum. However, when compared with Figure 17.21, it appears that the extra two ZFE pulses improve the reproduced speech more.

For a 20ms frame from the testfile BM2 the performance is highlighted in Figure 18.25. Observing Figure 18.25(b), it can be seen that the frequency spectrum changes from voiced to unvoiced at 900Hz. Furthermore, in the time domain it is difficult to determine the locations of the ZFE pulse.

The relative performances of the PWI-ZFE with 3-band MMBE and 13-band MMBE has been assessed through informal listening tests. Audibly the introduction of the extra frequency bands improves the natural quality of the speech signal. However, it

18.7. A HIGHER RATE 3.85KBPS MIXED-MULTIBAND EXCITATION SCHEME

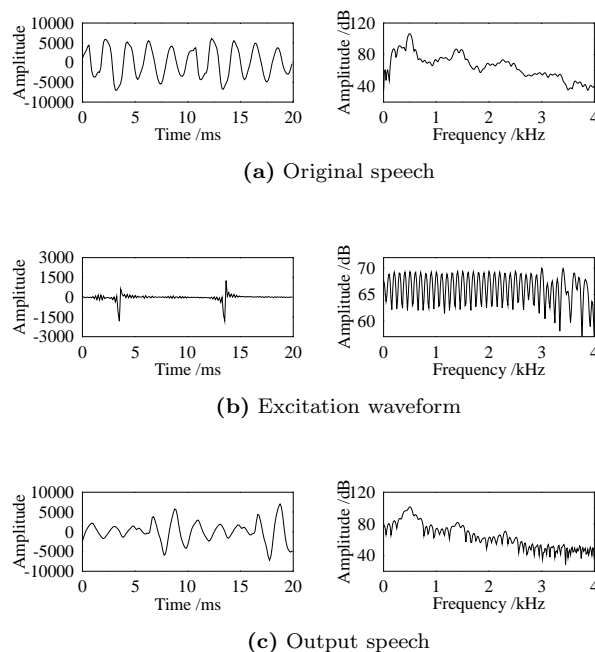


Figure 18.23: Time and frequency domain comparison of the a) original speech, b) **13-band MMBE ZFE** waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the mid vowel /ɜ:/ in the utterance ‘work’ for the testfile BM1. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

is debatable whether the improvement justifies the extra 1.5kbps bit rate contribution consumed by the extra bands. Through pairwise-comparison listening tests, detailed in Section 20.2 the 13-band MMBE extension to the PWI-ZFE speech coder performed better than the addition of two extra ZFE pulses. Given the problems with interpolation detailed in Section 17.11 this was to be expected. The conducted pairwise-comparison tests showed that 30.77% of listeners preferred the 13-band MMBE PWI-ZFE speech coder, with 5.13% of listeners preferring the 3-pulse PWI-ZFE scheme and 64.10% having no preference. Before offering our conclusions concerning this chapter, let us in the next section consider an interesting system design example, which is based on our previously designed 2.35 kbit/s speech codec.

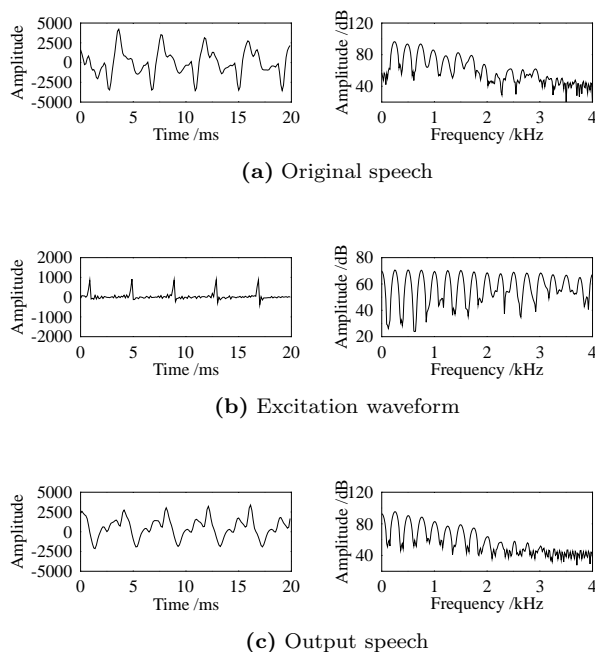


Figure 18.24: Time and frequency domain comparison of the a) original speech, b) **13-band MMBE ZFE** waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the liquid /r/ in the utterance '**rice**' for the testfile **BF2**. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

18.8 A 2.35 kbit/s Joint-detection based CDMA Speech Transceiver¹

18.8.1 Background

The standardisation of the third generation wireless systems has reached a mature state in Europe, the USA and Japan and the corresponding system developments are well under way right across the Globe. All three standard proposals are based on Wideband Code Division Multiple Access (W-CDMA), optionally supporting also joint multi-user detection in the up-link. In the field of speech and video source compression similarly impressive advances have been achieved and hence in this section a complete speech transceiver is proposed and its performance is quantified.

¹This section is based on F. C. A. Brooks, E. L. Kuan and L. Hanzo: A 2.35 kbit/s Joint-detection based CDMA Speech Transceiver; VTC'99, Houston, USA

18.8. A 2.35 KBIT/S JOINT-DETECTION CDMA SPEECH TRANSCEIVER 769

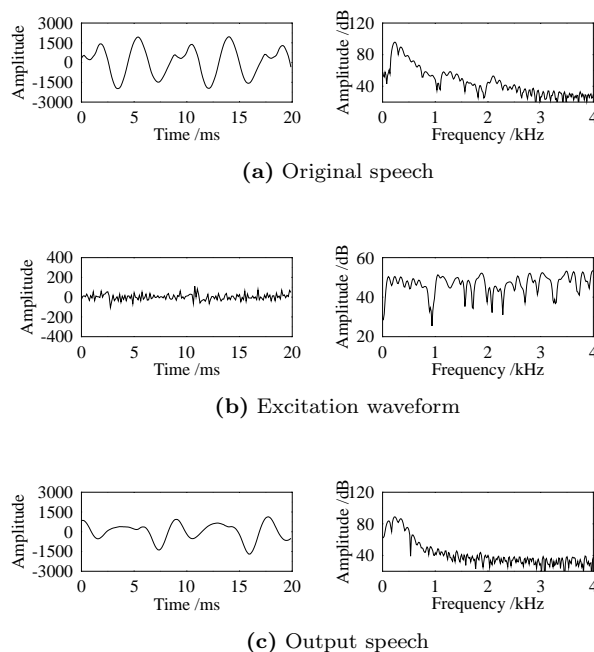


Figure 18.25: Time and frequency domain comparison of the a) original speech, b) **13-band MMBE ZFE** waveform and c) output speech after the pulse dispersion filter. The 20ms speech frame is the nasal /n/ in the utterance ‘**thrown for the testfile BM2**’. For comparison with the other coders developed in this study using the same speech segment please refer to Table 20.2.

18.8.2 The Speech Codec’s Bit Allocation

The codec’s bit allocation was summarized in Table 18.5, where again, 18 bits were reserved for LSF vector-quantization covering the groups of LSF parameters L0, L1, L2 and L3, where we used the nomenclature of the G.729 codec [269] for the groups of LSF parameters, since the G.729 codec’s LSF quantiser was used. A one-bit flag was used for the V/U classifier, while for unvoiced speech the RMS parameter was scalar quantized with 5-bits. For voiced speech the pitch-delay was restricted to $20 \rightarrow 147$ samples, thus requiring 7-bits for transmission. The ZFE amplitude parameters A and B were scalar quantized using 6-bits, since on the basis of our subjective and objective investigations we concluded that the 6-bit quantization constituted the best compromise in terms of bit rate and speech quality. The voicing strength for each frequency band was scalar quantized and since there were three frequency bands, a total of nine bits per 20 ms were allocated to voicing-strength quantisation. Thus the total number of bits for a 20ms frame became 26 or 47, yielding a transmission rate of 2.35kbps for the voice speech segments.

18.8.3 The Speech Codec's Error Sensitivity

Following the above description of the 2.35kbps speech codec we now investigate the extent of the reconstructed speech degradation inflicted by transmission errors. The error sensitivity is examined by individually corrupting each of the 47 bits detailed in Table 18.5 with a corruption probability of 10%. Employing a less than unity corruption probability is common practice, in order to allow the speech degradation caused by the previous corruption of a bit to decay, before the same bit is corrupted again, which emulates a practical transmission scenario realistically.

At the decoder for some of the transmitted parameters it is possible to invoke simple error checks and corrections. At the encoder isolated voiced, or unvoiced, frames are assumed to indicate a failure in the voiced-unvoiced decision and corrected, an identical process can be implemented at the decoder. For the pitch period parameter a smoothly evolving pitch track is created at the encoder by correcting any spurious pitch period values, and again, an identical process can be implemented at the decoder. Additionally, for voiced frame sequences phase continuity of the ZFE A and B amplitude parameters is maintained at the encoder, thus, if a phase change is perceived at the decoder, an error occurrence is assumed and the previous frame's parameters can be repeated.

Figure 18.26 displays the so-called Segmental Signal-to-Noise Ratio (SEGSNR) and cepstral distance (CD) objective speech measures for a mixture of male and female speakers, having British and American accents. Observing Figure 18.26 it can be seen that both the SEGSNR and CD objectives measures rate the error sensitivity of the different bits similarly. The most sensitive parameter is the voiced-unvoiced flag, followed closely by the pitch bits, while the least sensitive parameters are the three voicing strengths bits of the bands $B1 - B3$, as seen in Figure 18.26.

18.8.4 Channel Coding

In order to improve the performance of the system, channel coding was employed. Two types of error correction codes were used, namely, turbo codes and convolutional codes. Turbo coding is a powerful method of channel coding, which has been reported to produce excellent results [157, 325]. Convolutional codes were used as the component codes for the turbo coding and the coding rate was set to $r = 1/2$. We used a 7×7 block interleaver as the turbo interleaver. The FMA1 spread speech/data burst 1 [453] was altered slightly to fit the turbo interleaver. Specifically, the two data blocks were modified to transmit 25 data symbols in the first block and 24 symbols in the second one. In order to obtain the soft-decision inputs required by the turbo decoder, the Euclidean distance between the CDMA receiver's data estimates and each legitimate constellation point in the data modulation scheme was calculated. The set of distance values were then fed into the turbo decoder as soft inputs. The decoding algorithm used was the Soft Output Viterbi Algorithm (SOVA) [449, 450] with 8 iterations for turbo decoding. As a comparison, a half-rate, constraint-length three convolutional codec was used to produce a set of benchmark results. Note, however that while the turbo codec used so-called recursive systematic convolutional codecs, the convolutional codec was a non-recursive one, which has better distance properties.

18.8. A 2.35 KBIT/S JOINT-DETECTION CDMA SPEECH TRANSCEIVER 771

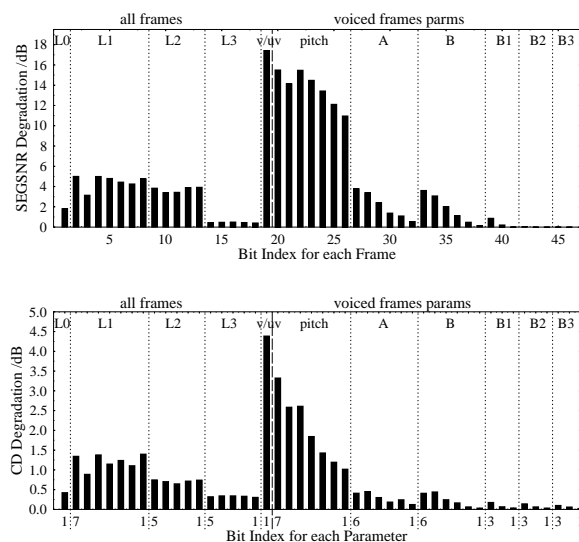


Figure 18.26: The error sensitivity of the different transmission bits for the 2.35kbps speech codec. For the CD degradation graph, containing the bit index for each parameter, bit 1 is the least significant bit.

18.8.5 The JD-CDMA Speech System

The JD-CDMA speech system used in our investigations is illustrated in Figure 18.27 for a two-user scenario. The encoded speech bits generated by the 2.35kbps prototype waveform interpolated (PWI) speech codec were channel encoded using a $\frac{1}{2}$ -rate turbo encoder having a frame length of 98 bits, including the convolutional codec's termination bits, where a 7×7 turbo interleaver was used. The encoded bits were then passed to a channel interleaver and modulated using 4-level Quadrature Amplitude Modulation (4-QAM). Subsequently, the modulated symbols were spread by the spreading sequence assigned to the user, where a random spreading sequence was used. The uplink conditions were investigated, where each user transmitted over a 7-path COST 207 Bad Urban channel [390], which is portrayed in Figure 18.28. Each path was faded independently using Rayleigh fading with a Doppler frequency of $f_D = 80$ Hz and a Baud rate of $R_b = 2.167$ MBaud. Variations due to path loss and shadowing were assumed to be eliminated by power control. The additive noise was assumed to be Gaussian with zero mean and a covariance matrix of $\sigma^2 \mathbf{I}$, where σ^2 is the variance of the noise. The burst structure used in our experiments mirrored the spread/speech burst structures of the FMA1 mode of the FRAMES proposal [453]. The Minimum Mean Squared Error Block Decision Feedback Equaliser (MMSE-BDFE) was used as the multiuser receiver [371], where perfect channel estimation and perfect decision feedback were assumed. The soft outputs for each user were obtained from the MMSE-BDFE and passed to the respective channel decoders. Finally, the decoded

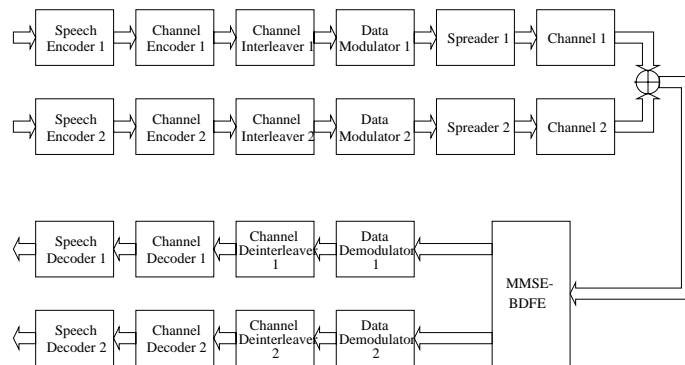


Figure 18.27: FRAMES-like two-user uplink CDMA system

bits were directed towards the speech decoder, where the original speech information was reconstructed.

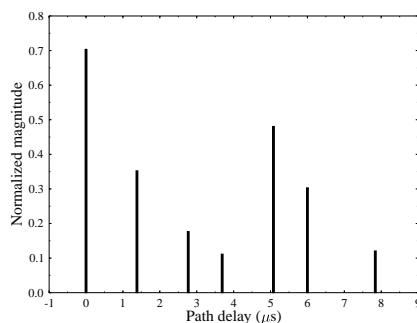


Figure 18.28: Normalized channel impulse response for a seven path Bad Urban channel [390].

18.8.6 System performance

The BER performance of the proposed system is presented in Figures 18.29 and 18.30. Specifically, Figure 18.29 portrays the BER performance of a two-user JD-CDMA speech transceiver. Three different sets of results were obtained for the uncoded, turbo-coded and non-systematic convolutional-coded systems, respectively. As it can be seen from the Figure, channel coding substantially improved the BER performance of the system. However, in comparing the BER performances of the turbo-coded system and the convolutional-coded system, convolutional coding appears to offer a slight performance improvement over turbo coding. This can be attributed to the fact

18.8. A 2.35 KBIT/S JOINT-DETECTION CDMA SPEECH TRANSCEIVER 773

that a short turbo interleaver was used, in order to maintain a low speech delay, while the non-systematic convolutional codec exhibited better distance properties. It is well-understood that turbo codecs achieve an improved performance in conjunction with long turbo interleavers. However, due to the low bit rate of the speech codec 47 bits per 20ms were generated and hence we were constrained to using a low interleaving depth for the channel codecs, resulting in a slightly superior convolutional coding performance.

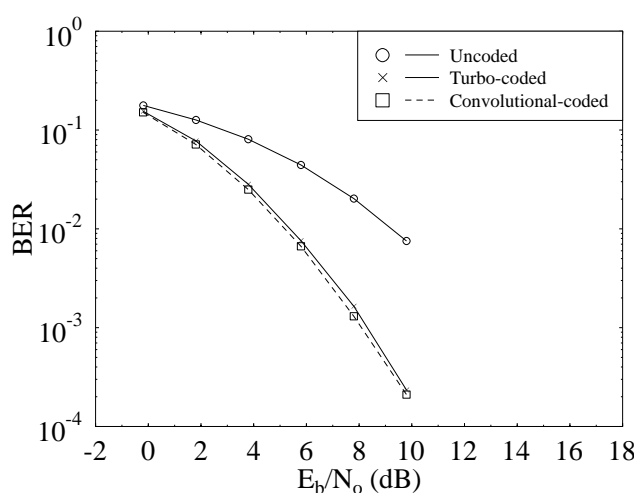


Figure 18.29: Comparison of the BER performance of an uncoded, convolutional-coded and turbo-coded two-user CDMA system, employing half-rate, constraint-length three constituent codes.

In Figure 18.30, the results were obtained by varying the number of users in the system between $K = 2$ and 6. The BER performance of the system degrades only slightly, when the number of users is increased. This is due to the employment of the joint detection receiver, which mitigates the effects of multiple access interference. It should also be noted that the performance of the system for $K = 1$ is also shown and the BER performances for $K = 2$ to 6 degrade only slightly from this single-user bound.

The SEGSR and CD objective speech measures for the decoded speech bits are depicted in Figure 18.31, where the turbo-coded and convolutional-coded systems were compared for $K = 2$ users. As expected on the basis of our BER curves, the convolutional codecs result in a lower speech quality degradation compared to the turbo codes, which were constrained to employ a low interleaver depth. Similar findings were observed in these Figures also for $K = 4$ and 6 users. Again, the speech performance of the system for different number of users is similar, demonstrating the efficiency of the JD-CDMA receiver.

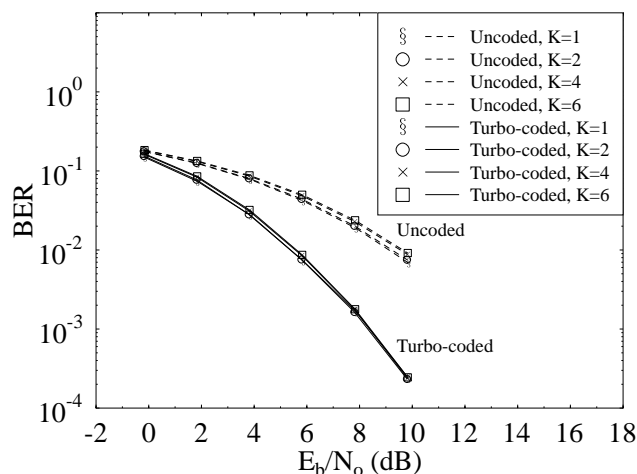


Figure 18.30: Comparison of the BER performance of an uncoded, convolutional-coded and turbo-coded CDMA system for $K = 2, 4$ and 6 users.

18.8.7 Conclusions on the JD-CDMA Speech Transceiver

The encoded speech bits generated by the 2.35kbps prototype waveform interpolated (PWI) speech codec were half-rate channel-coded and transmitted using a DS-CDMA scheme. At the receiver the MMSE-BDFE multiuser joint detector was used, in order to detect the information bits, which were then channel-decoded and passed on to the speech decoder. In our work, we compared the performance of turbo codes and convolutional codes. It was shown that the convolutional codes outperformed the more complex turbo codes in terms of their BER performance and also in speech SEGSR and CD degradation terms. This was due to the short interleaver constraint imposed by the low speech delay requirement, since turbo codes require a high interleaver length in order to perform effectively. It was also shown that the system performance was only slightly degraded, as the number of users was increased from $K = 2$ to 6 , demonstrating the efficiency of the JD-CDMA scheme.

18.9 Conclusion

This chapter has investigated the performance of MMBE when added to the LPC vocoder of Chapter 15 and the PWI-ZFE coder of Chapter 17. Initially, an overview of MBE was given, followed by detailed descriptions of the MMBE in both the encoder and decoder, given in Section 18.4 and 18.5, respectively.

Section 18.6.1 gave a detailed analysis of 2-band and 5-band MMBE added to the LPC vocoder, with Section 18.6.2 containing the analysis of 3-band MMBE added to

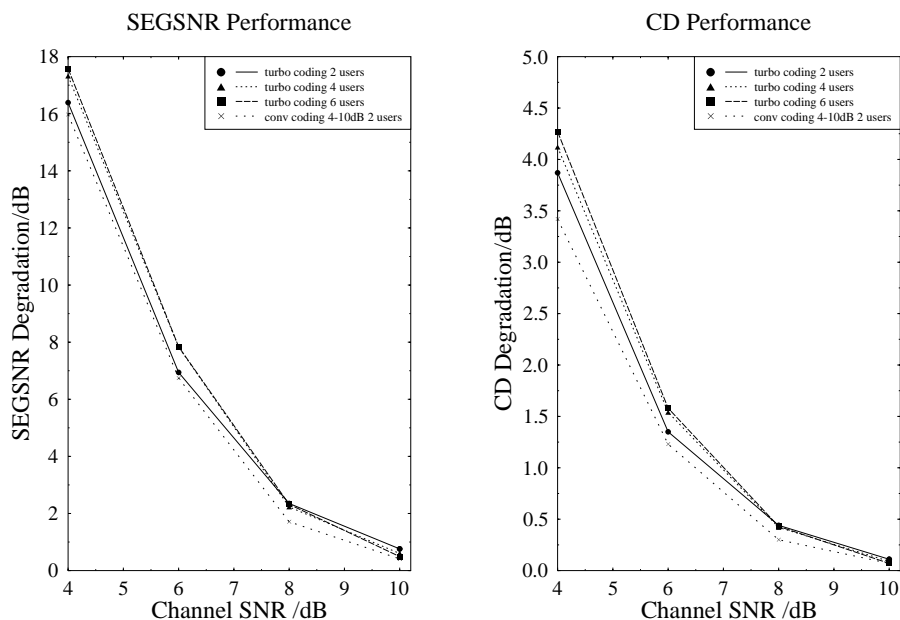


Figure 18.31: SEGSNR and CD objective speech measures for the decoded speech bits for $K = 2, 4$ and 6 users.

the PWI-ZFE coder. The 5-band MMBE LPC vocoder and the 3-band MMBE PWI-ZFE coder operated at similar bit rates, hence, they were compared through informal listening. It was found that the 3-band MMBE PWI-ZFE coder offered the best natural speech quality. The corresponding time- and frequency-domain waveforms of our coders investigated so far were summarized consistently using the same 20ms speech frames. The associated Figure numbers are detailed in Table 20.2.

Chapter 21

Comparison of Speech codecs and Transceivers

21.1 Background to Speech Quality Evaluation

The major difficulty associated with the assessment of speech quality is the consequence of a philosophical dilemma. Namely, should speech quality evaluation be based on unreliable, subjective human judgements or on reproducible objective evaluations, which may be highly uncorrelated with personal subjective quality assessments? Even high fidelity (HIFI) entertainment systems exhibit different subjective music reproduction qualities, let alone low-bit-rate speech codecs. It is practically impossible to select a generic set of objective measures in order to characterise speech quality, because all codecs result in different speech impairments. Some objective measures, which are appropriate for quantifying one type of distortion might be irrelevant to estimate another, just as one listener might prefer some imperfections to others. Using a statistically relevant, high number of trained listeners and various standardised tests mitigates the problems encountered, but incurs cost- and time penalties. During codec development usually quick and cost-efficient objective preference tests are used, followed by informal listening tests, before a full-scale formal subjective test is embarked upon.

The literature of speech quality assessment was documented in a range of excellent treatises by Kryter [469], Jayant and Noll [10], Kitawaki, Honda and Itoh [209,211].

In Reference [18] Papamichalis gives a comprehensive overview of the subject with references to Jayant's and Noll's work [10]. Further important contributions are due to Halka and Heute [470] as well as Wang, Sekey and Gersho [471].

21.2 Objective Speech Quality Measures

21.2.1 Introduction

Whether we evaluate the speech quality of a waveform codec, vocoder or hybrid codec, objective distance measures are needed to quantify the deviation of the codec's output signal from the input speech. In this respect any formal metric or distance measure of the mathematics, such as for example the Euclidean distance, could be employed to quantify the dissimilarity of the original and the processed speech signal, as long as symmetry, positive definiteness and the triangle inequality apply. These requirements were explicitly formulated as follows [210]:

- Symmetry: $d(x, y) = d(y, x)$,
- Positive Definiteness: $d(x, x) = 0$ and $d(x, y) > 0$, if $x \neq y$,
- Triangular Inequality: $d(x, y) \leq d(x, z) + d(y, z)$.

In practice the triangle inequality is not needed, but our distance measure should be easy to evaluate and preferably it ought to have some meaningful physical interpretation. The symmetry requires that there is no distinction between the reference signal and the speech to be evaluated in terms of distance. The positive definiteness implies that the distance is zero, if the reference and tested signals are identical.

A number of objective distance measures fulfill all criteria, some of which have waveform-related time-domain interpretations, while others have frequency-domain related physical meaning. Often time-domain waveform codecs such as eg, PCM are best characterised by the former ones, while frequency domain codecs, like transform and subband codecs by the latter. Analysis-by-synthesis hybrid codecs using perceptual error-weighting are the most difficult to characterise and usually only a combination of measures gives satisfactory results. Objective speech quality measures have been studied in depth by Quackenbush, Barnwell and Clements [21], hence here only a rudimentary overview is provided.

The simplest and most widely used metrics or objective speech quality measures are the signal-to-noise ratios (SNR), such as the conventional SNR, the segmental SNR (SEGSNR), and the frequency-weighted SNR [472]- [473]. Since they are essentially quantifying the waveform similarity of the original and the decoded signal, they are most useful in terms of evaluating waveform-coder distortions. Nonetheless, they are often invoked in medium-rate codecs, in order to compare different versions of the same codec, for example during the codec development process.

Frequency-domain codecs are often best characterised in terms of the spectral distortion between the original and processed speech signal, evaluating it either on the basis of the spectral fine structure, or - for example when judging the quality of a spectral envelope quantiser - in terms of the spectral envelope distortion. Some of the often used measures are the so-called spectral distance, log spectral distance, cepstral distance, log likelihood ratio, noise-masking ratios, and composite measures, most of which were proposed for example by Barnwell et al [472]- [474] during the late seventies and early eighties. However, most of the above measures are inadequate for quantifying the subjective quality of a wide range of speech-coder distortions. They

are particularly at fault predicting these quality degradations across different types of speech codecs. A particular deficiency of these measures is that when a range of different distortions are present simultaneously, these measure are incapable of evaluating the grade of the individual imperfections, although this would be desirable for codec developers.

Following the above introductory elaborations, let us now consider some of the widely used objective measures in a little more depth.

21.2.2 Signal to Noise Ratios

For discrete-time, zero-mean speech signals the error and signal energies of a block of N speech samples are given by:

$$E_e = \frac{1}{N} \sum_{u=1}^N (s(u) - \hat{s}(u))^2 \quad (21.1)$$

$$E_s = \frac{1}{N} \sum_{u=1}^N s^2(u). \quad (21.2)$$

Then the conventional Signal-to-Noise Ratio (SNR) is computed as:

$$SNR[dB] = 10 \log_{10} (E_s/E_e) \quad (21.3)$$

When computing the arithmetic means in Equations 21.1, the gross averaging over long sequences conceals the codecs' low SNR performance in low-energy speech segments and attributes unreasonably high objective scores to the speech codec. Computation of the geometric mean of the SNR guarantees higher correlation with perceptual judgements, because it gives proper weighting to the lower SNR performance in low-energy sections. This is achieved by computing the so-called segmental SNR (SEGSNR). Firstly, the speech signal is divided into segments of 10-20 ms and $SNR(u)$ [dB] is computed for $u = 1 \dots N$, ie for each segment in terms of dB. Then the segmental SNR(u) values are averaged in terms of dBs, as follows:

$$SEGSNR[dB] = \frac{1}{N} \sum_{n=1}^N SNR(u)[dB] \quad (21.4)$$

Equation 21.4 effectively averages the logarithms of the $SNR(u)$ values, which corresponds effectively to the computation of the geometric mean. This gives proper weighting to low-energy speech segments and therefore gives values more closely related to the subjective quality of the speech codec. A further refinement is to limit the segmental $SNR(u)$ terms to be in the range of $0 < SNR(u) < 40[dB]$, because outside this interval it becomes uncorrelated with subjective quality judgements.

21.2.3 Articulation Index

A useful frequency-domain related objective measure is the so-called articulation index (AI) proposed by Kryter in Reference [469]. The speech signal is split into 20 subbands of increasing bandwidths and the subband SNRs are computed. Their range is limited to $\text{SNR} = 30$ dB, and then the average SNR over the 20 bands is computed as follows:

$$AI = \frac{1}{20} \sum_{i=1}^{20} \text{SNR}_i \quad (21.5)$$

The subjective importance of the subbands is weighted by appropriately choosing the bandwidth of all subbands, which then contribute $1/20 - th$ of the total SNR. An important observation is that Kryter's original bandsplitting table stretches to 6100 Hz and when using a bandwidth of 4 kHz, the two top bands falling beyond 4 kHz are therefore neglected, limiting AI inherently to 90%. When using $B = 3\text{kHz}$, $AI \leq 80\%$. The evaluation of the AI is rather complex due to the bandsplitting operation.

21.2.4 Cepstral Distance

The cepstral distance (CD) is the most highly correlated objective measure, when compared to subjective measures. It maintains its high correlation over a wide range of codecs, speakers and distortions, while reasonably simple to evaluate. It is defined in terms of the cepstral coefficients of the reference and tested speech, as follows:

$$CD = \left[(c_0^{in} - c_0^{out})^2 + 2 \sum_{j=1}^{\infty} (c_j^{in} - c_j^{out})^2 \right]^{\frac{1}{2}} \quad (21.6)$$

The input and output cepstral coefficients are evaluated by the help of the linear predictive (LPC) filter coefficients a_j of the all-pole filter [210], which is elaborated on below.

Explicitly, the cepstral coefficients can be determined from the filter coefficients a_i ($i = 1 \dots p$) by the help of a recursive relationship, derived as follows. Let us denote the stable all-pole speech model by the polynomial $A(z)$ of order M in terms of z^{-1} , assuming that all its roots are inside the unit circle. It has been shown in Reference [475] that the following relationship holds for the Taylor series expansion of $\ln[A(z)]$:

$$\ln[A(z)] = - \sum_{k=1}^{\infty} c_k \cdot z^{-k}; \quad c_0 = \ln(E_p/R_0), \quad (21.7)$$

where the coefficients c_k are the cepstral coefficients and c_0 is the logarithmic ratio of the prediction error and the signal energy. By substituting

$$A(z) = 1 + \sum_{k=1}^{\infty} a_k \cdot z^{-k} \quad (21.8)$$

or by exploiting that $a_0 = 1$:

$$A(z) = 1 + \sum_{k=0}^M a_k \cdot z^{-k}. \quad (21.9)$$

Upon differentiating the left-hand side of Equation 21.7 with respect to z^{-1} we arrive at:

$$\frac{\delta[\ln A(z)]}{\delta z^{-1}} = \frac{1}{A(z)} \frac{\delta A(z)}{\delta z^{-1}} \quad (21.10)$$

$$\frac{\delta[\ln A(z)]}{\delta z^{-1}} = \frac{1}{\sum_{k=0}^M a_k \cdot z^{-k}} \sum_{k=1}^M k \cdot a_k \cdot z^{-(k-1)} \quad (21.11)$$

Differentiating the right-hand side of Equation 21.7 as well as equating it to the differentiated left-hand side according to Equation 21.9 yields:

$$\left(\sum_{k=0}^M a_k \cdot z^{-k}\right)^{-1} \sum_{k=1}^M k \cdot a_k \cdot z^{-(k-1)} = - \sum_{k=1}^{\infty} k \cdot c_k \cdot z^{-(k-1)} \quad (21.12)$$

Rearranging Equation 21.12 and multiplying both sides by z^{-1} results in Equation 21.13:

$$\sum_{k=1}^M k \cdot a_k \cdot z^{-k} = - \left(\sum_{k=0}^M a_k \cdot z^{-k}\right) \cdot \sum_{k=1}^{\infty} k \cdot c_k \cdot z^{-k}. \quad (21.13)$$

By expanding the indicated sums and performing the necessary multiplications the following recursive equations are resulted, which is demonstrated by an example in the next Section, Section 21.2.5:

$$c_1 = -a_1 \quad (21.14)$$

$$c_j = -\frac{1}{j} \left(j \cdot a_j + \sum_{i=1}^{j-1} i \cdot c_i \cdot a_{j-i} \right); \quad j = 2 \dots p \quad (21.15)$$

and by truncating the second sum in Equation 21.13 on the right-hand side at $2p$, since the higher order terms are of diminishing importance, we arrive at:

$$c_j = -\frac{1}{j} \sum_{k=1}^p (j-i) \cdot c_{j-i} \cdot a_i; \quad j = (p+1) \dots 2p. \quad (21.16)$$

Now, in possession of the filter coefficients the cepstral coefficients can be derived.

Having computed the cepstral coefficients $c_0 \dots c_{2p}$ we can determine the CD as repeated below for convenience:

$$CD = [(c_0^{in} - c_0^{out})^2 + 2 \cdot \sum_{j=1}^{2p} (c_j^{in} - c_j^{out})^2]^{\frac{1}{2}} \quad (21.17)$$

$$c_1 = a_1$$

$$\begin{aligned} c_j &= a_j - \sum_{r=1}^{j-1} \frac{r}{j} \cdot c_r \cdot a_{j-r} \quad \text{for } j = 2 \dots p \\ c_j &= - \sum_{r=1}^p \frac{j-r}{j} c_{j-r} \cdot a_r \quad \text{for } j = p+1, p+2, \dots, 3p \end{aligned} \quad (21.18)$$

where p is the order of the all-pole filter $A(z)$. The optimum predictor coefficients a_r are computed to minimise the energy of the prediction error residual:

$$e(u) = s(u) - \hat{s}(u) \quad (21.19)$$

This requires the solution of the following set of p equations:

$$\sum_{r=1}^p a_r \cdot R(|i-r|) = R(i) \quad \text{for } i = 1 \dots p, \quad (21.20)$$

where the autocorrelation coefficients are computed from the segmented and Hamming-Windowed speech, as follows. First the speech $s(u)$ is segmented into 20 ms or $N = 160$ samples long sequences. Then $s(u)$ is multiplied by the Hamming window function.

$$w(n) = 0.54 - 0.45 \cos \frac{2\pi n}{N} \quad (21.21)$$

In order to smooth the frequency domain oscillations introduced by the rectangular windowing of $s(u)$. Now the autocorrelation coefficients $R(i)$ $i = 1 \dots p$ are computed from the windowed speech $s_w(n)$ as

$$R(i) = \sum_{n=0}^{N-1-i} s_w(n) \cdot s_w(n+i) \quad i = 1 \dots p \quad (21.22)$$

Finally, Equation 21.20 is solved for the predictor coefficients $a(i)$ by the Levinson-Durbin algorithm [182]:

$$\begin{aligned} E(0) &= R(0) \\ \epsilon_i &= \left[\sum_{j=1}^{i-1} a_j^{(i-1)} \cdot R(i-j) \right] / E^{(i-1)} \quad i = 1 \dots p \\ a_{(i)}^{(i)} &= r_i \\ a_j &= a_j^{(i-1)} - k_i \cdot a_{i-j}^{(i-1)} \quad j = 1 \dots (i-1) \\ E^{(i)} &= (1 - \epsilon_i^2) \end{aligned} \quad (21.23)$$

where r_i , $i = 1 \dots p$ are the reflection coefficients. After p iterations ($i = 1 \dots p$) the set of LPC coefficients is given by:

$$a_j = a_j^{(p)} \quad j = 1 \dots p. \quad (21.24)$$

and the prediction gain is given by $G = E(i)/E^{10}$. The computation of the CD is summarised in the flow chart of Figure 21.1:

It is plausible that the CD measure is a spectral domain parameter, since it is related to the LPC filter coefficients of the speech spectral envelope. In harmony with our expectations, it is shown in Reference [210] that the CD is identical to the logarithmic root mean square spectral distance (LRMS-SD) between the input and output spectral envelopes often used in speech quality evaluations:

$$LRMS - SD = \left[\int_{-\pi}^{\pi} |\ln|b_{in}/A_{in}(f)|^2 - \ln|G_{out}/A_{out}(f)|^2| df \right]^{\frac{1}{2}}. \quad (21.25)$$

In the next Section we will consider a simple example.

21.2.5 Example: Computation of Cepstral Coefficients

Let us make the derivation of Equations 21.14-21.16 plausible by expanding the sums in Equation 21.13 and by computing the multiplications indicated. For the sake of this let us assume $p = 4$ and compute $c_1 \dots c_{2p}$:

$$\begin{aligned} a_1 z^{-1} + 2a_2 z^{-2} + 3a_3 z^{-3} + 4a_4 z^{-4} &= \\ &= -(c_1 z^{-1} + 2c_2 z^{-2} + 3c_3 z^{-3} + 4c_4 z^{-4} + 5c_5 z^{-5} + 6c_6 z^{-6} \\ &\quad + 7c_7 z^{-7} + 8c_8 z^{-8}) \cdot (1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} + a_4 z^{-4}) \end{aligned} \quad (21.26)$$

By computing the product at the right-hand side, we arrive at:

$$\begin{aligned} a_1 z^{-1} + 2a_2 z^{-2} + 3a_3 z^{-3} + 4a_4 z^{-4} &= \\ &= c_1 z^{-1} + 2c_2 z^{-2} + 3c_3 z^{-3} + 4c_4 z^{-4} + 5c_5 z^{-5} + 6c_6 z^{-6} \\ &\quad + 7c_7 z^{-7} + 8c_8 z^{-8} + c_1 a_1 z^{-2} + 2c_2 a_1 z^{-3} + 3c_3 a_1 z^{-4} \\ &\quad + 4c_4 a_1 z^{-5} + 5c_5 a_1 z^{-6} + 6c_6 a_1 z^{-7} + 7c_7 a_1 z^{-8} + 8c_8 a_1 z^{-9} \\ &\quad + c_1 a_2 z^{-3} + 2c_2 a_2 z^{-4} + 3c_3 a_2 z^{-5} + 4c_4 a_2 z^{-6} + 5c_5 a_2 z^{-7} \\ &\quad + 6c_6 a_2 z^{-8} + 7c_7 a_2 z^{-9} + 8c_8 a_2 z^{-10} + c_1 a_3 z^{-4} + 2c_2 a_3 z^{-5} \\ &\quad + 3c_3 a_3 z^{-6} + 4c_4 a_3 z^{-7} + 5c_5 a_3 z^{-8} + 6c_6 a_3 z^{-9} \\ &\quad + 7c_7 a_3 z^{-10} + 8c_8 a_3 z^{-11} \end{aligned} \quad (21.27)$$

Now by matching the terms of equal order in z^{-1} on both sides:

z^{-1} :

$$c_1 = -a_1 \quad (21.28)$$

z^{-2} :

$$\begin{aligned} 2a_2 &= 2c_2 + a_1 c_1 \\ 2c_2 &= -a_1 c_1 - 2a_2 \end{aligned} \quad (21.29)$$

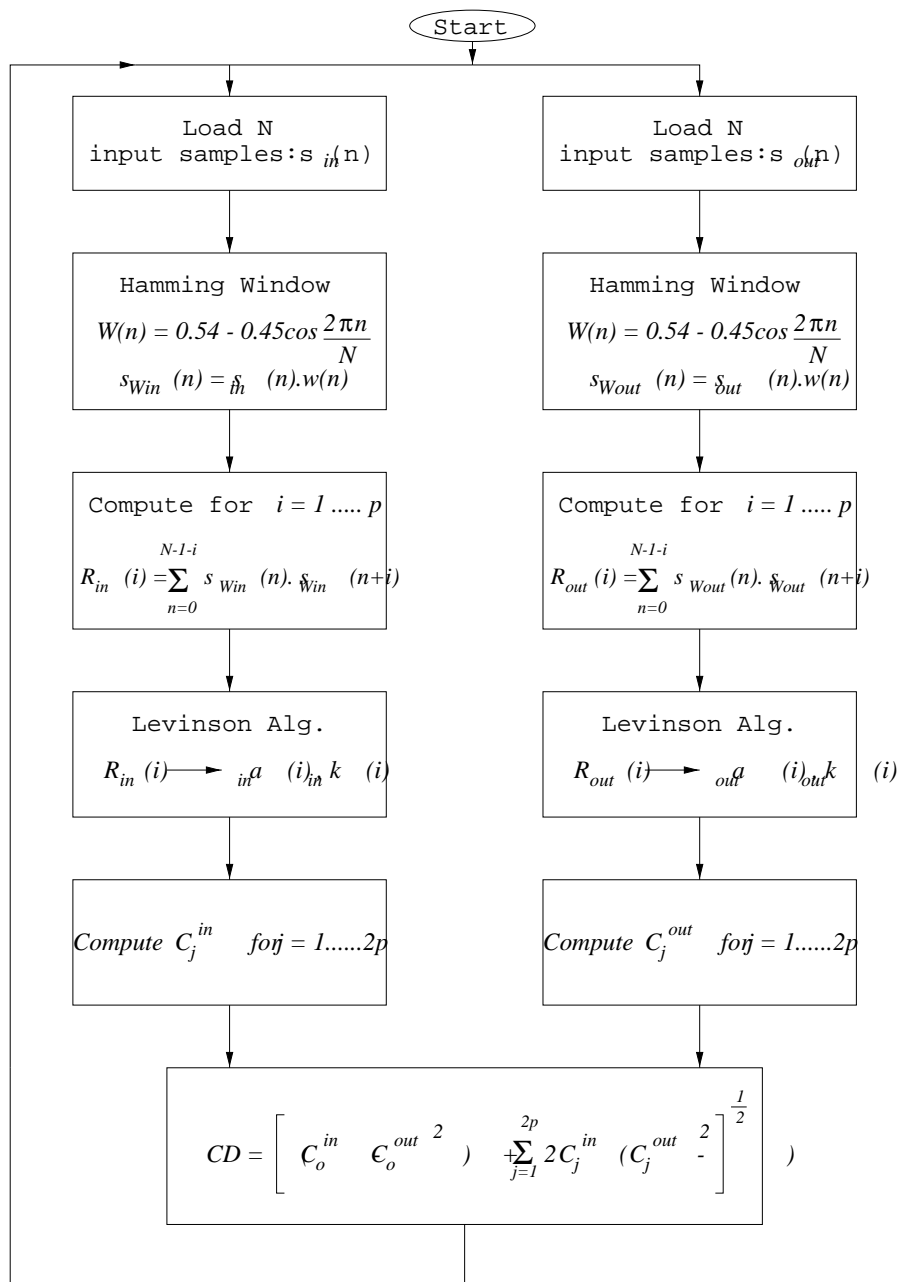


Figure 21.1: Cepstrum Distance Computation Flowchart

$$z^{-3}; \quad 3c_3 = -3a_3 - 2a_1c_2 - a_2c_1 \quad (21.30)$$

$$z^{-4}; \quad 4c_4 = -4a_4 - 3a_1c_3 - 2a_2c_2 \quad (21.31)$$

In general:

$$jc_j = -ja_j - \sum_{i=1}^{j-1} ic_ia_{j-i}; \quad j = 1 \dots p. \quad (21.32)$$

But also, there exists a number of terms with an order of higher than p that must cancel each other on the right-hand side of 21.27: z^{-5} :

$$5c_5 + 4c_4a_1 + 3c_3a_2 + 2c_2a_3 + c_1a_4 = 0 \quad (21.33)$$

$$5c_5 = -4c_4a_1 - 3c_3a_2 - 2c_2a_3 - c_1a_4 \quad (21.34)$$

$$z^{-6}; \quad 6c_6 = -5c_5a_1 - 4c_4a_2 - 3c_3a_3 - 2c_2a_4 \quad (21.35)$$

$$z^{-7}; \quad 7c_7 = -6c_6a_1 - 5c_5a_2 - 4c_4a_3 - 3c_3a_4 \quad (21.36)$$

$$z^{-8}; \quad 8c_8 = -7c_7a_1 - 6c_6a_2 - 5c_5a_3 - 4c_4a_4 \quad (21.37)$$

In general:

$$jc_j = -\sum_{i=1}^p (j-i)c_{j-i}a_i; \quad j = p+1 \dots \quad (21.38)$$

Let us now continue our review of various objective speech quality measures in the spirit of Papamichalis' discussions [18] in the next Section.

21.2.6 Logarithmic likelihood ratio

The likelihood ratio (LR) distance measure introduced by Itakura uses also the LPC coefficients of the input and output spectral envelope to quantify the spectral deviation introduced by the speech codec. The LR is defined as the ratio of the LPC residual energy before and after speech coding. Since the LPC coefficients $\underline{a}_{rin} = [a_0, a_1, \dots, a_p]$ are computed by Durbin's algorithm to minimise the LPC residual's energy, replacing \underline{a}_{rin} by another LPC coefficient vector \underline{a}_{r1} out computed from the decoded speech certainly increases the LPC residual energy, therefore $LR \geq 1$.

The formal definition of the LR is given by

$$LR = \frac{\underline{a}_{out}^T \underline{R}^{out} \underline{a}_{out}}{\underline{a}_{in}^T \underline{R}^{in} \underline{a}_{in}} \quad (21.39)$$

where $\underline{a}_{in}, \underline{R}^{in}$ and $\underline{a}_{out}, \underline{R}^{out}$ represent the LPC filter coefficient vectors and auto-correlation matrices of the input as well as output speech, respectively. The LR ?? defined in (15) is non-symmetric, which contradicts to our initial requirements.

Fortunately, this can be rectified by the symmetric transformation:

$$LRS = \frac{LR + 1/LR}{2} - 1 \quad (21.40)$$

Finally, the symmetric logarithmic LR (SLLR) is computed from:

$$SLLR = 10 \log_{10}(LRS) \quad (21.41)$$

The computational complexity incurred is significantly reduced, if LR is evaluated instead of the matrix multiplications required by (15) exploiting the following relationship:

$$\underline{a}^T \cdot \underline{Ra} = R_a(0)R(0) + 2 \sum_{i=1}^P R_a(i) \cdot R(i), \quad (21.42)$$

where $R(i)$ and $R_a(i)$ represent the autocorrelation coefficients of the signal and that of the LPC filter coefficients \underline{a} , as computed in (4).

21.2.7 Euclidean Distance

If any comprehensive set of spectral parameters closely related to the spectral deviation between input and output speech is available, the Euclidean instance between the sets of input and output speech parameters gives useful insights into the distortions inflicted. Potentially suitable sets are the LPC coefficients, the reflection coefficients computed in (12), the autocorrelation coefficients given in (11), the so called line spectrum frequencies (LSF) most often used recently or the highly robust logarithmic area ratios (LAR). LARs are defined as

$$LAR_i = \ln \frac{1 + r_i}{1 - r_i} \quad i = 1 \dots p \quad (21.43)$$

are very robust against channel errors and have a fairly limited dynamic range, which alleviate their quantisation. With this definition of LARs the Euclidean distance is formulated as:

$$D_{LAR} = \left[\sum_{i=1}^p (LAR_i^{in} - LAR_i^{out})^2 \right]^{\frac{1}{2}} \quad (21.44)$$

21.3 Subjective Measures [18]

Once the development of a speech codec is finalised, objective and informal subjective tests are followed by formal subjective tests. Depending on the type, bitrate and quality of the specific codec different subjective tests are required to test quality and intelligibility. Quality is usually tested by the so called Diagnostic Acceptability Measure (DAM), paired preference tests or the most wide-spread Mean Opinion Score (MOS). Intelligibility is tested by Consonant-Vowel-Consonant (CVC) logatours or by Dynamic Rhythm Tests (DRT). Formal subjective speech assessment is generally

Speech Impairment	Typical of
Fluttering	Amplitude Modulated Speech
Thin	Highpass filtered Speech
Rasping	Peak Clipped Speech
Muffled	Lowpass filtered speech
Interrupted	Packetised Speech
Nasal	Low-bit-rate Vocoders

Table 21.1: Typical terms to characterise speech impairments in DAM tests ©Papamichalis [18], 1987

Background	Typical of
Hissing	Noisy speech
Buzzing	Tandemed Dig Systems
Babbling	Low-bit-rate codecs with bit errors
Rumbling	Low-frequency-noise marked speech

Table 21.2: Typical terms for background qualities in DAM tests ©Papamichalis [18], 1987

a lengthy investigation carried out by specially trained unbiased crew using semi-standardised test material, equipment and conditions.

21.3.1 Quality Tests

In Diagnostic Acceptability Measure tests the trained listener is asked to rate the speech codec tested using phonetically balanced sentences from the so-called Harvard list in terms of both speech quality and background quality. Some terms used at Dynastat (USA) to describe speech imperfections are listed following Papamichalis in Table 21.3.1 [18]. As regards to background qualities the sort of terms used at Dynastat were summarised following Papamichalis in Table 21.3.1 [18]: The speech and background qualities are rates in the listed categories on a 100 point scale by each listener and then their average scores are evaluated for each category, giving also the standard deviations and standard errors. Before averaging the results of various categories appropriate weighting factors can be used to emphasise features particularly important for a specific application of the codec.

In *Pair-wise preference tests* the listeners compare always the same sentence processed by two different codecs, even if a high number of codecs has to be tested. To ensure consistency in the preferences also unprocessed and identically processed sentences can be included. The results are summarised in the preference matrix. If the comparisons show a clean preference order for differently processed speech and an approximately random preference (50%) for identical codecs in the preference matrix's main diagonal, the results are accepted. However, if no clear preference order is established, different tests have to be deployed.

21.4 Comparison of Subjective and Objective Measures

21.4.1 Background

An interesting comparison of the objective articulation index (AI) described in Section 21.2.3 and of various subjective tests was given by Kryter [476], as shown in Figure 21.3. Observe that the lower the size of the test vocabulary used, the higher are the intelligibility scores for a fixed AI value, which is due to the less subtle differences inherent in a smaller test vocabulary.

The *Modulated noise reference unit (MNRU)* proposed by Law and Seymour [288] to relate subjective quality to objective measures is widely used by the CCITT as well. The MNRU block-diagram is shown in Figure 21.2.

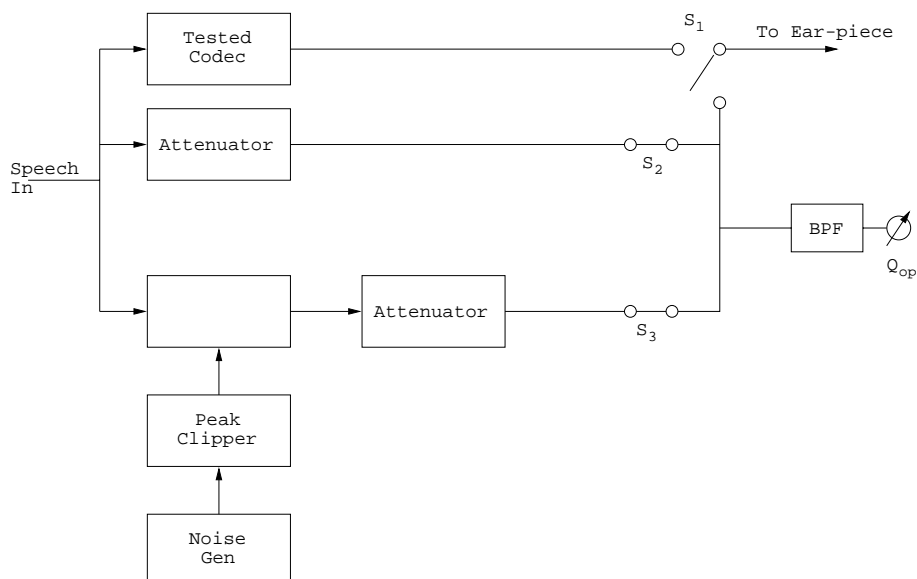


Figure 21.2: Modulated noise reference unit block diagram

The MNRU is used to add noise, amplitude modulated by the speech test material, to the reference speech signal, rendering the noise speech-correlated. The SNR of the reference signal is gradually lowered by the listener using the attenuators in Figure 21.2 to perceive identical loudness and subjective qualities, when comparing the noisy reference signal and the tested codec's output speech. During this adjustment and comparison phase the switches 52 and 53 are closed and 51 is being switched between the reference and tested speech signals. Once both speech signals make identical subjective impressions, switches 52 and 53 are used to measure the reference signal's and noise signal's power and hence the so-called opinion equivalent Q [dB] (Q_{op}) expressed in terms of the SNR computed. Although the Q_{op} [dB] value appears to be an objective measured value, it depends on various listeners subjective judgements

and therefore is classified as a subjective measure. The Q_{op} [dB] value is easily translated into the more easily interpreted MOS measure using the reference speech's MOS vs. Q_{op} characteristic depicted in Figure 21.3.

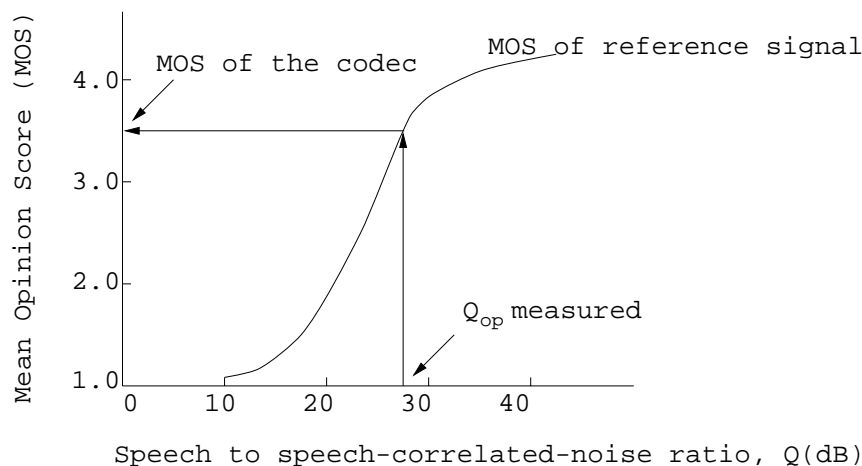


Figure 21.3: Translating Q_{op} into MOS

21.4.2 Intelligibility tests

In intelligibility tests the listeners are asked to recognise which one of a pair of words is uttered, where the two words only differ in one phoneme, which is a consonant [477]. Alternatively, consonant-vowel-consonant (CVC) logatours can also be used. According to Papamichalis in the so-called diagnostic rhyme test (DRT) developed by Dynastat (Texas, USA) [18] a set of 96 rhyming pairs of words are utilised, some of which are: meat-beat, pear-tear, saw-thaw, bond-pond, etc. The pairs are specially selected to test the following phonetic attributes: voicing, nasality, sustention, sibilantion, graveness and compactness. If, for example, the codec under test consistently fails to distinguish between vast-fast, zoo-sue, goat-coat, ie, to deliver clear voiced sounds such as v, z, g etc, it points out for the designer that the codec's long-term predictor responsible for the spectral fine-structure, or voicing information in the spectrum does not work properly. Consistently grouping and evaluating the recognition failures vital information can be gained about the codecs shortcomings. Typical DRT values are between 75 and 95 and for high intelligibility DRT ≥ 90 is required.

In a similar fashion, most objective and subjective measures can be statistically related to each other, but the goodness of match predicted for new codecs varies over a wide range. For low-bit-rate codecs one of the most pertinent relationships devised is [211]:

$$MOS = 0.04CD_2 - 0.80CD + 3.565 \quad (21.45)$$

This formula is the best second order fit to a high number of MOS-CD measurements carried out over a variety of codecs and imperfections.

In summary, speech quality evaluation is usually based on quick objective assessments during codec development, followed by extensive formal subjective tests, when the development is finalised. A range of objective and subjective measures was described, where the most popular objective measures are the simple time-domain SEG SNR [dB] and the somewhat more complex, frequency-domain CD [dB] measure. The CD objective measure is deemed to have the highest correlation with the most widely applicable subjective measure, the MOS, and their relationship is expressed in Equation 21.45. Having reviewed a variety of objective and subjective speech quality measures, let us now compare a range of previously considered speech codecs in the next Section.

21.5 Subjective Speech Quality of Various Codecs

In previous Chapters we have characterised many different speech codecs. Here we attempt a rudimentary comparison of some of the previously described codec schemes in terms of their subjective and objective speech quality as well as error sensitivity. We will conclude this Chapter by incorporating some of the codecs concerned in various wireless transceivers and portray their SEG SNR versus channel SNR performance. Here we refer back to Figure 5.6 and with reference to Cox's work [1, 2] we populate this Figure with actual formally evaluated Mean Opinion Score (MOS) values, which are shown in Figure 21.4. Observe 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 8kbps G.729 codec is the most recent addition to this range of ITU standard schemes, which significantly outperforms all previous standard ITU codecs in robustness terms. The performance target of the 4kbps ITU codec (ITU4) is also to maintain this impressive set of specifications. The family of codecs, which were designed for various mobile radio systems, such as the 13kbps RPE GSM scheme, the 7.95kbps IS-54, the IS-96, the 6.7kbps JDC and 3.45kbps 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 subjective speech quality of a range of speech codecs is characterised in Figure 21.4. While during our introductory discussions we portrayed the waveform coding, vocoding and hybrid coding families in a similar, but more inaccurate, stylised illustration, this Figure is based on large-scale formal comparative studies.

The 2.4 kbps Federal Standard codec 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 ADPCM schemes are waveform codecs. They exhibit a low implementational complexity associated with a modest bitrate economy. The remaining codecs belong to the hybrid coding family and achieve significant bitrate economies at the cost of increased complexity and delay.

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. This activity overlapped with the G.723.1 developments. The 6.4 kbps mode maintains a speech quality similar to the G.711, G.726, G.727, G.728 and G.728 codecs, while the 5.3 mode exhibits a speech quality similar to the cellular speech codecs of the late eighties. Work is under way at the time of writing towards the standardisation of a 4 kbps ITU scheme, which we refer to here as ITU4.

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 CELP-based codecs. Observe in the Figure that there is also an identical-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 8 kbps 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-Synchronous Innovation CELP (PSI-CELP) scheme. The IS-96 American CDMA system also has its own standardised CELP-based speech codec, which is a variable-rate scheme, allowing 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 21.4. 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.

Following the above elaborations as regards to the perceived speech quality of a range of speech codecs, let us now consider their objective speech quality and robustness aspects in the next Section.

21.6 Error Sensitivity Comparison of Various Codecs

As a rudimentary objective speech quality measure based bit-sensitivity comparison, in Figure 21.5 we portrayed the SEGSNR degradations of a number of speech codecs for a range of bit error rates (BER), when applying random errors. The SEGSNR degradation is in general not a reliable measure of speech quality, nonetheless, it indicates adequately, how rapidly this objective speech quality measure decays for the various codecs, when exposed to a given fixed BER. As expected, the backwards-adaptive G.728 and the forward-adaptive G.723.1 schemes, which have been designed mainly for benign wireline connections, have the fastest SGSNR degradation upon increasing the BER. By far the best performance is exhibited by the G.729 scheme, followed by the 13 kbps GSM codec. In the next Section we will highlight, how these codecs perform over Gaussian and Rayleigh-fading channels using three different

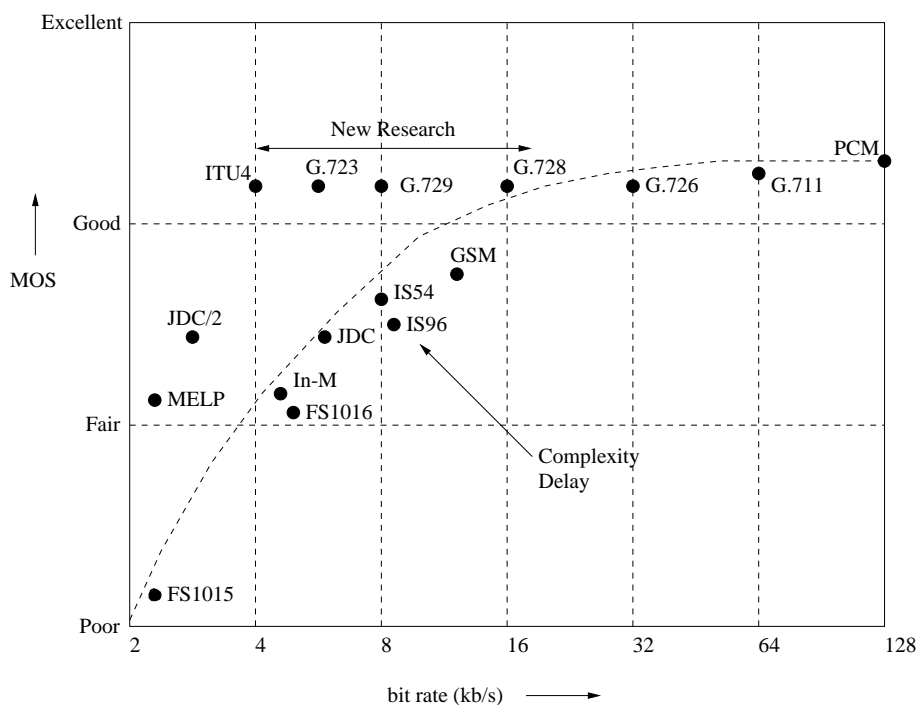


Figure 21.4: Subjective speech quality of various codecs [1] ©IEEE, 1996

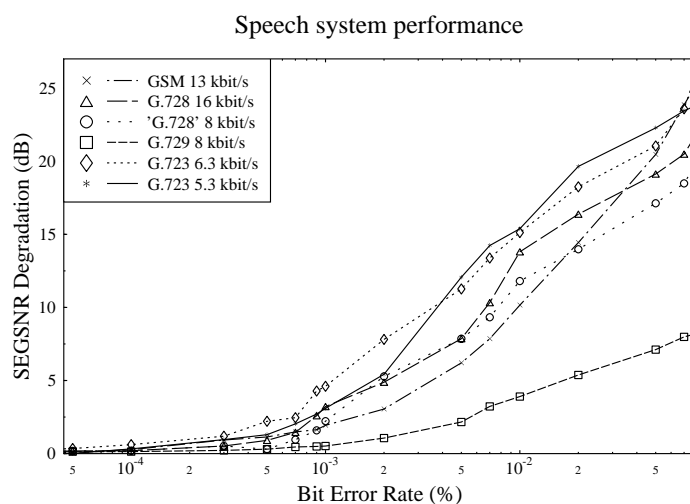


Figure 21.5: SEGSNR degradation versus BER for the investigated speech codecs

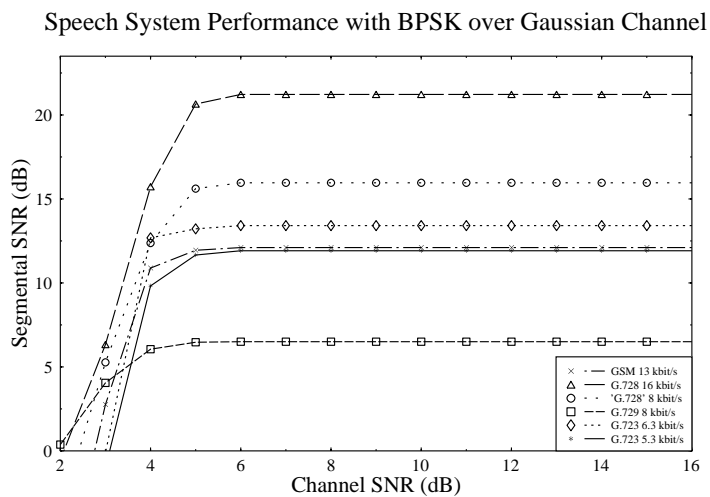


Figure 21.6: SEGSNR versus channel SNR performance of various speech codecs using the BCH(254,130,18) code and BPSK over Gaussian channels

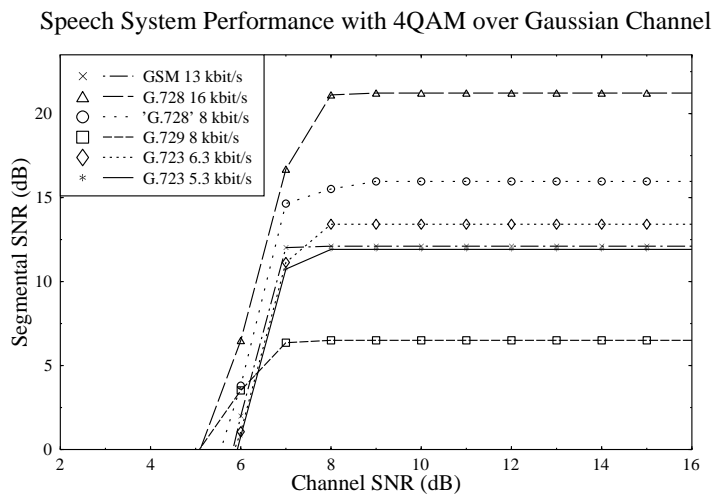


Figure 21.7: SEGSNR versus channel SNR performance of various speech codecs using the BCH(254,130,18) code and 4QAM over Gaussian channels

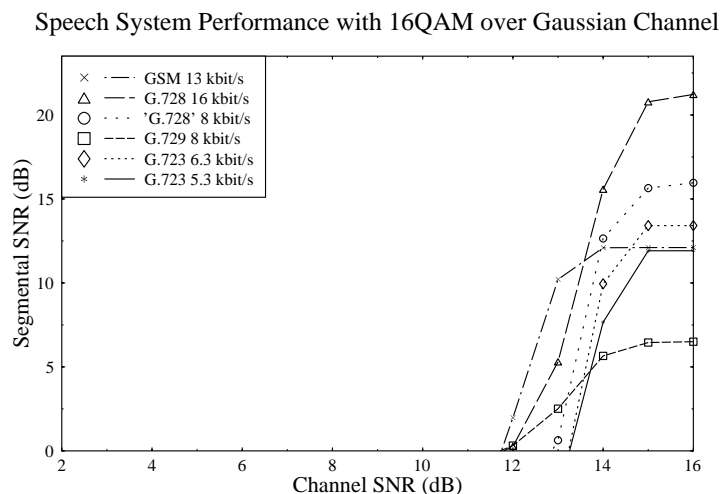


Figure 21.8: SEGSR versus channel SNR performance of various speech codecs using the BCH(254,130,18) code and 16QAM over Gaussian channels

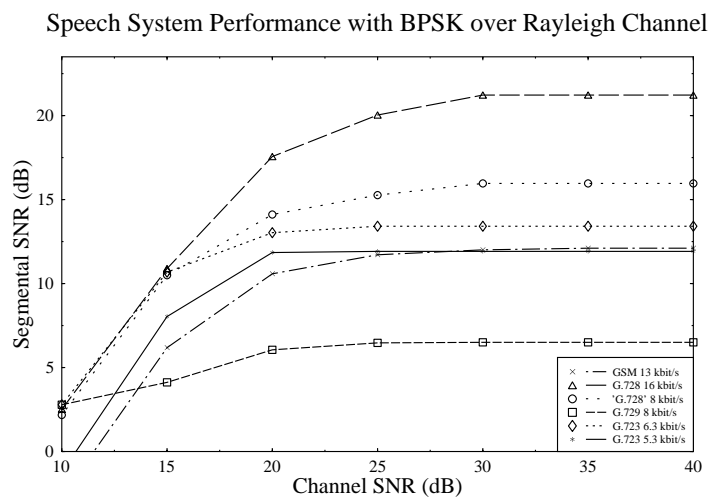


Figure 21.9: SEGSR versus channel SNR performance of various speech codecs using the BCH(254,130,18) code and BPSK over Rayleigh channels

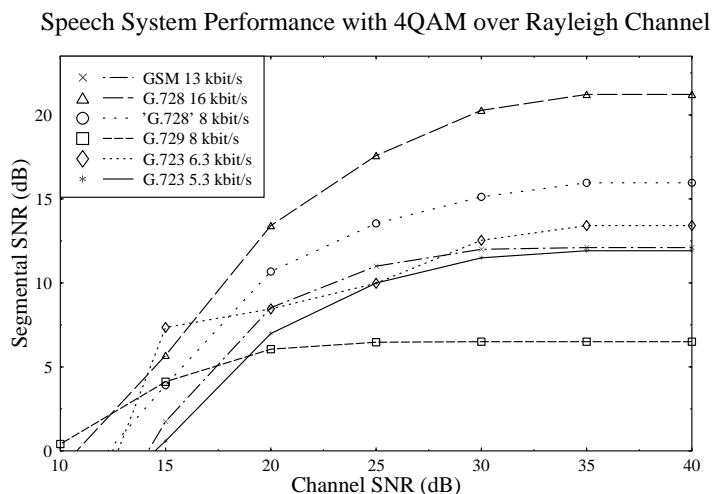


Figure 21.10: SEGSR versus channel SNR performance of various speech codecs using the BCH(254,130,18) code and 4QAM over Rayleigh channels

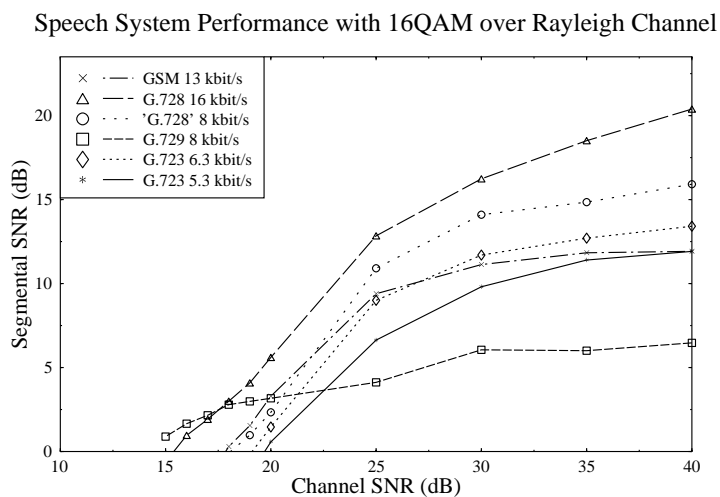


Figure 21.11: SEGSR versus channel SNR performance of various speech codecs using the BCH(254,130,18) code and 16QAM over Rayleigh channels

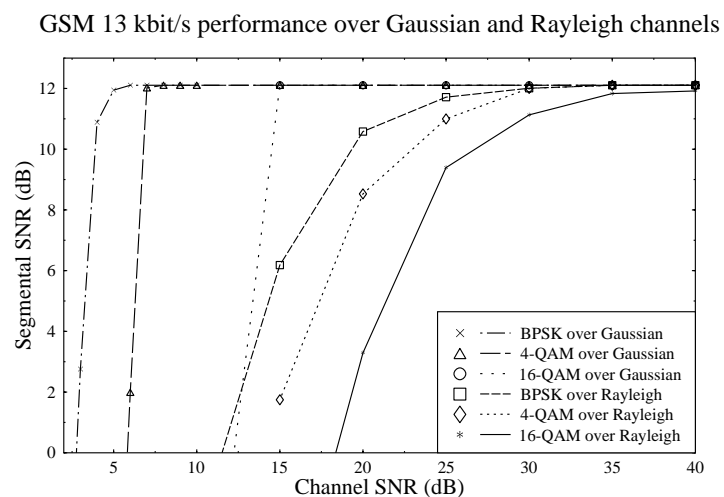


Figure 21.12: SEGSNR degradation versus channel SNR performance of the 13 kbps RPE-LTP GSM speech codec using the BCH(254,130,18) code and BPSK, 4QAM as well as 16QAM over both Gaussian and Rayleigh channels

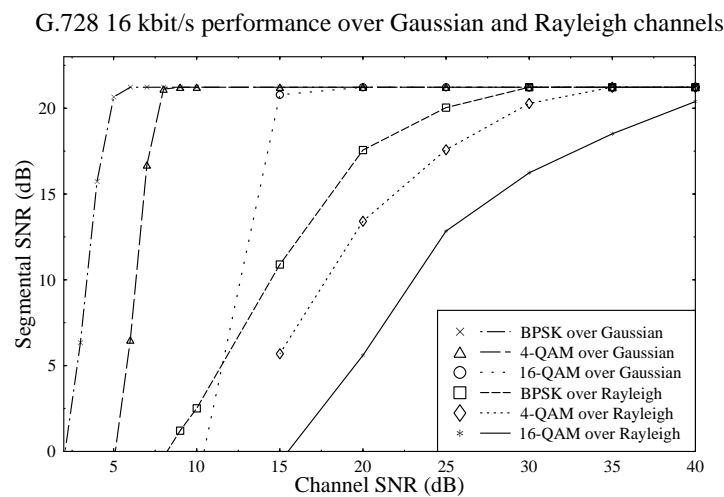


Figure 21.13: SEGSNR degradation versus channel SNR performance of the 16 kbps backward-adaptive G.728 speech codec using the BCH(254,130,18) code and BPSK, 4QAM as well as 16QAM over both Gaussian and Rayleigh channels

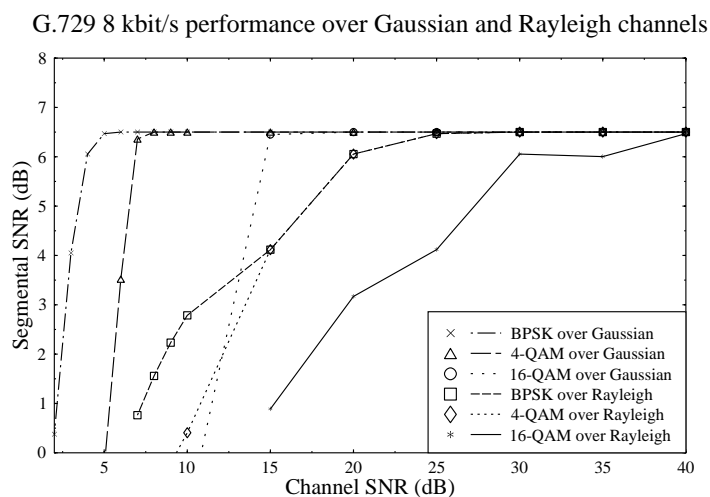


Figure 21.14: SEGSR degradation versus channel SNR performance of the 8 kbps forward-adaptive G.729 speech codec using the BCH(254,130,18) code and BPSK, 4QAM as well as 16QAM over both Gaussian and Rayleigh channels

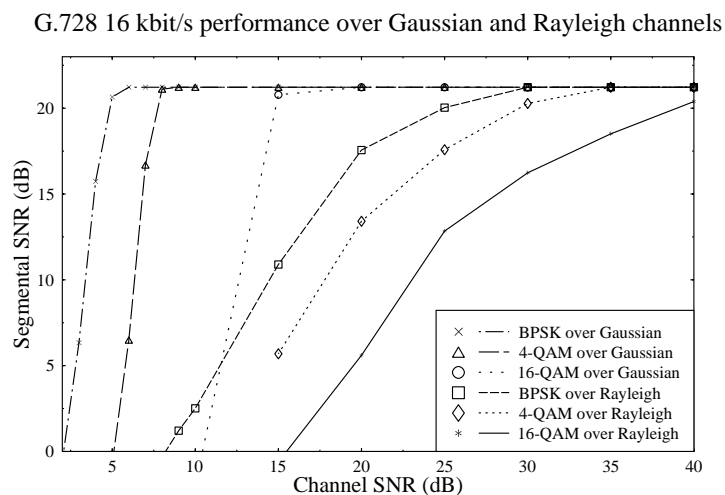


Figure 21.15: SEGSR degradation versus channel SNR performance of the 5.3 kbps G.723.1 speech codec using the BCH(254,130,18) code and BPSK, 4QAM as well as 16QAM over both Gaussian and Rayleigh channels

Codec	Rate (kbps)	BPSK		4-QAM		16-QAM	
		AWGN	Ray.	AWGN	Ray.	AWGN	Ray.
GSM	13	4	20	7	27	13	34
G.728	16	5	26	8	30	15	40
'G.728'	8	5	25	7	31	15	35
G.729	8	4	19	7	20	14	28
G.723.1	6.4	4	18	8	31	15	35
G.723.1	5.3	4	19	7	29	15	35

Table 21.3: Minimum required channel SNR for maintaining less than 1 dB SEGSNR degradation for the investigated speech transceivers using the BCH(254,130,18) code and BPSK, 4QAM as well as 16QAM over both Gaussian and Rayleigh channels

transceivers.

21.7 Objective Speech-performance of Various Transceivers

In this Section we embarked on the comparison of the previously analysed speech codecs under identical experimental circumstances, when used in identical transceivers over both Gaussian and Rayleigh channels. These results are portrayed in Figures 21.6-21.11, which will be detailed during our further discourse. Three different modems, namely 1, 2 and 4 bits/symbol Binary Phase Shift Keying (BPSK), 4-level Quadrature Amplitude Modulation (4QAM) and 16-QAM were employed in conjunction with the six different modes of operations of the four speech codecs that were protected by the BCH(254,130,18) channel codec. Note that here no specific source-sensitivity matched multi-class channel coding was invoked, in order to ensure identical experimental conditions for all speech codecs. Although in general the SEGSNR is not a good absolute measure, when comparing speech codecs operating on the basis of different coding algorithms, it can be used as a robustness-indicator, exhibiting a decaying characteristic for degrading channel conditions and hence allowing us to identify the minimum required channel SNRs for the various speech codecs and transceiver modes. Hence here we opted for using the SEGSNR in these comparisons, providing us with an opportunity to point out its weaknesses on the basis of our apriori knowledge as regards to the codecs' formally established subjective quality.

Under error-free transmission and no background-noise conditions the subjective speech quality of the 16 kbps G.728 scheme, the 8 kbps G.729 codec and the 6.4 kbps G.723.1 arrangement is characterised by a Mean Opinion Score (MOS) of approximately four. In other words, their perceived speech quality is quite similar, despite their different bitrates. Their similar speech quality at such different bitrates is a ramification of the fact that they represent different milestones during the evolution of speech codecs, since they were contrived in the above chronological order. They also exhibit different implementational complexities. The 13 kbps GSM codec and the 5.3 kbps G.723 arrangements are slightly inferior in terms of their subjective quality, both of which are characterised by an MOS of about 3.5. We note here, however that

there exists a recently standardised so-called enhanced full-rate, 13 kbps GSM speech codec, which also has an MOS of about four under perfect channel conditions.

The above subjective speech qualities are not reflected by the corresponding SEGSNR curves portrayed in Figures 21.6-21.11. For example, the 8 kbps G.729 codec has the lowest SEGSNR, although it has an MOS similar to G.728 and the 6.4 kbps G.723.1 schemes in terms of subjective speech quality. As expected, this is due to the high-pass filtering operation at its input, as well as a ramification of the more pronounced perceptually motivated speech quality optimisation, as opposed to advocating high-quality wave-form reproduction. A further interesting comparison is offered by the 8 kbps 'G.728-like' non-standard codec, which exhibits a higher SEGSNR than the identical bitrate G.729 scheme, but sounds significantly inferior to the G.729 arrangement. These differences become even more conspicuous, when they are exposed to channel errors in the low-SNR region of the curves. In terms of error resilience the G.729 scheme is far the best in the group of codecs tested. The minimum required channel SNR values for the various transceivers over the Gaussian and Rayleigh channels are summarised in Table 21.3. Observe in the Rayleigh-channel curves of Figures 21.9-21.11 that the backwards-adaptive codecs have a rapidly decaying performance curve, whereas for example the G.729 forward-adaptive ACELP scheme exhibits a more robust behaviour. Lastly, in Figures 21.12-21.15 we organised our previous results in a different way, plotting all the different SEGSNR versus channel SNR curves related to a specific speech codec in the same Figure, allowing a direct comparison of the expected speech performance of the various transceivers over various channel conditions.

Bibliography

- [1] R. V. Cox and P. Kroon, "Low bit-rate speech coders for multimedia communications," *IEEE Comms. Mag.*, pp. 34–41, December 1996.
- [2] R. V. Cox, "Speech coding and synthesis," in *Speech coding standards* (W. Kleijn and K. Paliwal, eds.), ch. 2, pp. 49–78, Elsevier, 1995.
- [3] R. Steele, *Delta modulation systems*. Pentech Press, London, 1975.
- [4] K. Cattermole, *Principles of Pulse Code Modulation*. London: Hiffe Books, 1969.
- [5] J. Markel and A. Gray, Jr., *Linear Prediction of Speech*. New York: Springer-Verlag, 1976.
- [6] L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*. Prentice-Hall, 1978.
- [7] B. Lindblom and S. Ohman, *Frontiers of Speech Communication Research*. Academic Press, 1979.
- [8] J. V. Tobias, ed., *Foundations of Modern Auditory Theory*. NY, U.S.A: Academic Press, 1970. ISBN: 0126919011.
- [9] B. S. Atal and J. R. 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* [499], pp. 614–617.
- [10] N. Jayant and P. Noll, *Digital coding of waveforms, Principles and applications to speech and video*. 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 *Proc. 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. R. Schroeder and B. S. 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*. Prentice-Hall Englewood Cliffs, New Jersey, 1987.
- [19] J. Deller, J. Proakis, and J. Hansen, *Discrete-time processing of speech signals*. Prentice-Hall, 1987.
- [20] P. Lieberman and S. Blumstein, *Speech physiology, speech perception, and acoustic phonetics*. Cambridge University Press, 1988.
- [21] S. Quackenbush, T. Barnwell III, and M. Clements, *Objective measures of speech quality*. Prentice Hall, Englewood Cliffs, NJ, 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 Proc.*, 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 Tech. 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 Tech. Journal*, vol. 63, pp. 587–626, April 1984.

BIBLIOGRAPHY

863

- [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 Tech. Journal*, vol. 63, pp. 2193–2213, December 1984.
- [27] J. Hagenauer, "Source-controlled channel decoding," *IEEE Transactions on Communications*, vol. 43, pp. 2449–2457, Sept 1995.
- [28] B. S. Atal, V. Cuperman, and A. Gersho, eds., *Advances in Speech Coding*. Kluwer Academic Publishers, Jan 1991. ISBN: 0792390911.
- [29] A. Ince, ed., *Digital Speech Processing: Speech Coding, Synthesis and Recognition*. Kluwer Academic Publishers, 1992.
- [30] J. Anderson and S. Mohan, *Source and Channel Coding - An Algorithmic Approach*. Kluwer Academic Publishers, 1993.
- [31] A. Kondozi, *Digital Speech: Coding for low bit rate communications systems*. John Wiley, 1994.
- [32] W. Keijn and K. Paliwal, eds., *Speech Coding and Synthesis*. Elsevier Science, 1995.
- [33] W. C. Jakes, ed., *Microwave Mobile Communications*. John Wiley and Sons, 1974. ISBN 0-471-43720-4.
- [34] R. Steele and L. Hanzo, eds., *Mobile Radio Communications*. IEEE Press-John Wiley, 2 ed., 1999.
- [35] R. Steele, "Towards a high capacity digital cellular mobile radio system," *Proc. of the IEE*, vol. 132, Part F, pp. 405–415, August 1985.
- [36] R. Steele and V. Prabhu, "High-user density digital cellular mobile radio system," *IEE Proc.*, vol. 132, Part F, pp. 396–404, August 1985.
- [37] R. Steele, "The cellular environment of lightweight hand-held portables," *IEEE Communications Magazine*, pp. 20–29, July 1989.
- [38] L. Hanzo and J. Stefanov, "The Pan-European Digital Cellular Mobile Radio System – known as GSM," in Steele [180], ch. 8, pp. 677–765.
- [39] J. D. Gibson, ed., *The Mobile Communications Handbook*. CRC Press and IEEE Press, 1996.
- [40] W. Lee, *Mobile cellular communications*. New York: McGraw Hill, 1989.
- [41] J. Parsons and J. Gardiner, *Mobile communication systems*. London: Blackie, 1989.
- [42] D. Parsons, *The mobile radio propagation channel*. London: Pentech Press, 1992.

- [43] D. Greenwood and L. Hanzo, "Characterisation of mobile radio channels," in Steele [180], ch. 2, pp. 92–185.
- [44] R. Edwards and J. Durkin, "Computer prediction of service area for VHF mobile radio networks," *Proc IRE* 116 (9), pp. 1493–1500, 1969.
- [45] M. Hata, "Empirical formula for propagation loss in land mobile radio," *IEEE Trans. on Vehicular Technology*, vol. 29, pp. 317–325, August 1980.
- [46] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, "Field strength and its variability in VHF and UHF land mobile service," *Review of the Electrical Communication Laboratory*, vol. 16, pp. 825–873, September–October 1968.
- [47] W.T.Webb, "Sizing up the microcell for mobile radio communications," *IEE Electronics and communications Journal*, vol. 5, pp. 133–140, June 1993.
- [48] J. G. Proakis, *Digital Communications*. McGraw Hill, 3rd ed., 1995.
- [49] C. Shannon, *Mathematical Theory of Communication*. University of Illinois Press, 1963.
- [50] J. Hagenauer, "Quellengesteuerte kanalcodierung fuer sprach- und tonuebertragung im mobilfunk," *Aachener Kolloquium Signaltheorie*, pp. 67–76, 23–25 March 1994.
- [51] A. J. Viterbi, "Wireless digital communications: A view based on three lessons learned," *IEEE Communications Magazine*, pp. 33–36, September 1991.
- [52] L. Hanzo and J. P. Woodard, "An intelligent multimode voice communications system for indoor communications," *IEEE Transactions on Vehicular Technology*, vol. 44, pp. 735–748, Nov 1995. ISSN 0018-9545.
- [53] L. Hanzo, R. A. 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.
- [54] K. H. H. Wong and L. Hanzo, "Channel coding," in Steele [180], ch. 4, pp. 347–488.
- [55] R. A. Salami, L. Hanzo, R. Steele, K. H. J. Wong, and I. Wassell, "Speech coding," in Steele [180], ch. 3, pp. 186–346.
- [56] "Special issue: The European Path Toward UMTS," *IEEE Personal Communications: The magazine of nomadic communications and computing*, vol. 2, Feb 1995.
- [57] European Commission, *Advanced Communications Technologies and Services (ACTS)*, Aug 1994. Workplan DGXIII-B-RA946043-WP.
- [58] Telcomm. Industry Association (TIA), Washington, DC, *Dual-mode subscriber equipment - Network equipment compatibility specification, Interim Standard IS-54*, 1989.

BIBLIOGRAPHY

865

- [59] R. Prasad, *CDMA for Wireless Personal Communications*. Artech House, May 1996. ISBN 0890065713.
- [60] Telcomm. Industry Association (TIA), Washington, DC, *Mobile station - Base station compatibility standard for dual-mode wideband spread spectrum cellular system, EIA/TIA Interim Standard IS-95*, 1993.
- [61] 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.
- [62] A. Urie, M. Streeton, and C. Mourot, "An advanced TDMA mobile access system for UMTS," *IEEE Comms. Mag.*, pp. 38–47, February 1995.
- [63] "European RACE D731 public deliverable," September 1995. Mobile communication networks, general aspects and evolution.
- [64] Research and Development Centre for Radio Systems, Japan, *Public Digital Cellular (PDC) Standard, RCR STD-27*.
- [65] "Feature topic: Software Radios," *IEEE Communications Magazine*, vol. 33, pp. 24–68, May 1995.
- [66] G. D. Forney Jr, R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE Journal on Selected Areas in Communications*, vol. 2, pp. 632–647, Sept 1984.
- [67] K. Feher, "Modems for emerging digital cellular mobile systems," *IEEE Tr. on VT*, vol. 40, pp. 355–365, May 1991.
- [68] W. Webb, L. Hanzo, and R. Steele, "Bandwidth-efficient QAM schemes for rayleigh-fading channels," *IEE Proceedings*, vol. 138, pp. 169–175, June 1991.
- [69] A. Wright and W. Durtler, "Experimental performance of an adaptive digital linearized power amplifier," *IEEE Tr. on VT*, vol. 41, pp. 395–400, November 1992.
- [70] R. Wilkinson et al, "Linear transmitter design for MSAT terminals," in *Proc. of 2nd Int. Mobile Satellite Conference*, June 1990.
- [71] P. Kenington, R. Wilkinson, and J. Marvill, "Broadband linear amplifier design for a PCN base-station," in *Proceedings of IEEE Vehicular Technology Conference (VTC'91)* [490], pp. 155–160.
- [72] S. Stapleton and F. Costescu, "An adaptive predistorter for a power amplifier based on adjacent channel emissions," *IEEE Tr. on VT*, vol. 41, pp. 49–57, February 1992.
- [73] S. Stapleton, G. Kandola, and J. Cavers, "Simulation and analysis of an adaptive predistorter utilizing a complex spectral convolution," *IEEE Tr. on VT*, vol. 41, pp. 387–394, November 1992.

- [74] 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, July 15–28 1995.
- [75] J. M. Torrance and L. Hanzo, "Latency considerations for adaptive modulation in a slow Rayleigh fading channel," in *Proceedings of IEEE VTC '97* [487], pp. 1204–1209.
- [76] H. Nyquist, "Certain factors affecting telegraph speed," *Bell System Tech Jnl*, p. 617, April 1928.
- [77] W. T. Webb and L. Hanzo, *Modern Quadrature Amplitude Modulation: Principles and Applications for Wireless Communications*. IEEE Press-Pentech Press, 1994. ISBN 0-7273-1701-6.
- [78] H. R. Raemer, *Statistical communication theory and applications*. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1969.
- [79] K. Feher, ed., *Digital communications - satellite/earth station engineering*. Prentice Hall, 1983.
- [80] Y. C. Chow, A. R. Nix, and J. P. McGeehan, "Analysis of 16-APSK modulation in AWGN and rayleigh fading channel," *Electronic Letters*, vol. 28, pp. 1608–1610, November 1992.
- [81] B. Sklar, *Digital communications - Fundamentals and Applications*. Prentice Hall, 1988.
- [82] J. Torrance, "Digital modulation," phd mini-thesis, Dept. of Electronics and Computer Science, Univ. of Southampton, UK, 1996.
- [83] J. Cavers, "An analysis of pilot symbol assisted modulation for rayleigh fading channels," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 686–693, Nov 1991.
- [84] F. Adachi, "Error rate analysis of differentially encoded and detected 16APSK under rician fading," *IEEE Tr. on Veh. Techn.*, vol. 45, pp. 1–12, February 1996.
- [85] J. McGeehan and A. Bateman, "Phase-locked transparent tone in band (TTIB): A new spectrum configuration particularly suited to the transmission of data over SSB mobile radio networks," *IEEE Transactions on Communications*, vol. COM-32, no. 1, pp. 81–87, 1984.
- [86] A. Bateman, "Feedforward transparent tone in band: Its implementation and applications," *IEEE Trans. Veh. Tech.*, vol. 39, pp. 235–243, August 1990.
- [87] J. M. Torrance and L. Hanzo, "Comparative study of pilot symbol assisted modem schemes," in *Proceedings of IEE Conference on Radio Receivers and Associated Systems (RRAS'95)* [486], pp. 36–41.

BIBLIOGRAPHY

867

- [88] M. L. Moher and J. H. Lodge, "TCMP – a modulation and coding strategy for rician fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 1347–1355, December 1989.
- [89] S. Sampei and T. Sunaga, "Rayleigh fading compensation method for 16-QAM in digital land mobile radio channels," in *Proceedings of IEEE Vehicular Technology Conference (VTC'89)*, (San Francisco, CA, USA), pp. 640–646, IEEE, 1–3 May 1989.
- [90] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 1991.
- [91] A. Bateman and J. McGeehan, "Feedforward transparent tone in band for rapid fading protection in multipath fading," in *IEE Int. Conf. Comms.*, vol. 68, pp. 9–13, 1986.
- [92] J. Cavers, "The performance of phase locked transparent tone in band with symmetric phase detection," *IEEE Trans. on Comms.*, vol. 39, pp. 1389–1399, September 1991.
- [93] R. Steele and W. Webb, "Variable rate QAM for data transmission over Rayleigh fading channels," in *Proceedings of Wireless '91*, (Calgary, Alberta), pp. 1–14, IEEE, 1991.
- [94] W. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Transactions on Communications*, vol. 43, no. 7, pp. 2223–2230, 1995.
- [95] 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* [483], pp. 709–713.
- [96] S. Chua and A. Goldsmith, "Variable-rate variable-power mQAM for fading channels," in *Proceedings of IEEE VTC '96* [484], pp. 815–819.
- [97] A. Goldsmith and S. Chua, "Adaptive coded modulation for fading channels," *IEEE Tr. on Communications*, vol. 46, pp. 595–602, May 1998.
- [98] D. A. Pearce, A. G. Burr, and T. C. Tozer, "Comparison of counter-measures against slow Rayleigh fading for TDMA systems," in *IEE Colloquium on Advanced TDMA Techniques and Applications*, (London, UK), pp. 9/1–9/6, IEE, 28 October 1996. digest 1996/234.
- [99] W. C. Y. Lee, "Estimate of channel capacity in Rayleigh fading environment," *IEEE Trans. on Vehicular Technology*, vol. 39, pp. 187–189, Aug 1990.
- [100] N. Morinaga, "Advanced wireless communication technologies for achieving high-speed mobile radios," *IEICE Transactions on Communications*, vol. 78, no. 8, pp. 1089–1094, 1995.
- [101] J. M. 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.

- [102] J. M. Torrance and L. Hanzo, "Optimisation of switching levels for adaptive modulation in a slow Rayleigh fading channel," *Electronics Letters*, vol. 32, pp. 1167–1169, 20 June 1996.
- [103] J. M. Torrance and L. Hanzo, "Demodulation level selection in adaptive modulation," *Electronics Letters*, vol. 32, pp. 1751–1752, 12 September 1996.
- [104] J. Torrance and L. Hanzo, "Performance upper bound of adaptive QAM in slow Rayleigh-fading environments," in *Proc. of IEEE ICCS'96 / ISPACS'96*, (Singapore), pp. 1653–1657, IEEE, 25–29 November 1996.
- [105] J. Torrance and L. Hanzo, "Adaptive modulation in a slow Rayleigh fading channel," in *Proc. of IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC'96)*, vol. 2, (Taipei, Taiwan), pp. 497–501, IEEE, 15–18 October 1996.
- [106] A. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. on Communications*, vol. 45, pp. 1218–1230, Oct. 1997.
- [107] M.-S. Alouini and A. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," to appear *IEEE Tr. on Veh. Techn.*, 1999. <http://www.systems.caltech.edu>.
- [108] A. Goldsmith, "The capacity of downlink fading channels with variable rate and power," *IEEE Tr. on Veh. Techn.*, vol. 46, pp. 569–580, Aug. 1997.
- [109] A. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Tr. on Inf. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.
- [110] J. Woodard and L. Hanzo, "A low delay multimode speech terminal," in *Proceedings of IEEE VTC '96* [484], pp. 213–217.
- [111] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. Cambridge University Press, 1992.
- [112] C. Wong and L. Hanzo, "Upper-bound of a wideband burst-by-burst adaptive modem," in *Proceeding of VTC'99 (Spring)* [479].
- [113] 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.
- [114] T. Liew, C. Wong, and L. Hanzo, "Block turbo coded burst-by-burst adaptive modems," in *Proceeding of Microcoll'99, Budapest, Hungary*, pp. 59–62, 21–24 March 1999.
- [115] C. Wong, T. Liew, and L. Hanzo, "Blind modem mode detection aided block turbo coded burst-by-burst wideband adaptive modulation," in *Proceeding of ACTS Mobile Communication Summit '99* [478].
- [116] K. Narayanan and L. Cimini, "Equalizer adaptation algorithms for high speed wireless communications," in *Proceedings of IEEE VTC '96* [484], pp. 681–685.

BIBLIOGRAPHY

869

- [117] J. Wu and A. H. Aghvami, "A new adaptive equalizer with channel estimator for mobile radio communications," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 467–474, August 1996.
- [118] Y. Gu and T. Le-Ngoc, "Adaptive combined DFE/MLSE techniques for ISI channels," *IEEE Transactions on Communications*, vol. 44, pp. 847–857, July 1996.
- [119] A. Clark and R. Harun, "Assessment of kalman-filter channel estimators for an HF radio link," *IEE Proceedings*, vol. 133, pp. 513–521, Oct 1986.
- [120] R. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *BSTJ*, vol. 46, pp. 1775–1796, December 1966.
- [121] M. Zimmermann and A. Kirsch, "The AN/GSC-10/KATHRYN/ variable rate data modem for HF radio," *IEEE Trans. Commun. Techn.*, vol. CCM-15, pp. 197–205, April 1967.
- [122] E. Powers and M. Zimmermann, "A digital implementation of a multichannel data modem," in *Proc. of the IEEE Int. Conf. on Commun.*, (Philadelphia, USA), 1968.
- [123] B. Saltzberg, "Performance of an efficient parallel data transmission system," *IEEE Trans. Commun. Techn.*, pp. 805–813, December 1967.
- [124] R. Chang and R. Gibby, "A theoretical study of performance of an orthogonal multiplexing data transmission scheme," *IEEE Trans. Commun. Techn.*, vol. COM-16, pp. 529–540, August 1968.
- [125] S. Weinstein and P. Ebert, "Data transmission by frequency division multiplexing using the discrete fourier transform," *IEEE Trans. Commun. Techn.*, vol. COM-19, pp. 628–634, October 1971.
- [126] Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'80* [488], pp. 964–967.
- [127] B. Hirosaki, "An orthogonally multiplexed QAM system using the discrete fourier transform," *IEEE Trans. Commun.*, vol. COM-29, pp. 983–989, July 1981.
- [128] L. J. 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.
- [129] K. Kammeyer, U. Tuisel, H. Schulze, and H. Bochmann, "Digital multicarrier transmission of audio signals over mobile radio channels," *European Transactions on Telecommunications*, vol. 3, pp. 243–253, May–Jun 1992.
- [130] F. Mueller-Roemer, "Directions in audio broadcasting," *Jnl Audio Eng. Soc.*, vol. 41, pp. 158–173, March 1993.

- [131] G. Plenge, "DAB - a new radio broadcasting system - state of development and ways for its introduction," *Rundfunktech. Mitt.*, vol. 35, no. 2, 1991.
- [132] M. Alard and R. Lassalle, "Principles of modulation and channel coding for digital broadcasting for mobile receivers," *EBU Review, Technical No. 224*, pp. 47–69, August 1987.
- [133] *Proc. 1st Int. Symp., DAB*, (Montreux, Switzerland), June 1992.
- [134] I. Kalet, "The multitone channel," *IEEE Tran. on Comms*, vol. 37, pp. 119–124, February 1989.
- [135] H. Kolb, "Untersuchungen über ein digitales mehrfrequenzverfahren zur datenübertragung," in *Ausgewählte Arbeiten über Nachrichtensysteme*, no. 50, Universität Erlangen-Nürnberg, 1982.
- [136] H. Schüssler, "Ein digitales Mehrfrequenzverfahren zur Datenübertragung," in *Professoren-Konferenz, Stand und Entwicklungsaussichten der Daten und Telekommunikation*, (Darmstadt, Germany), pp. 179–196, 1983.
- [137] K. Preuss, "Ein Parallelverfahren zur schnellen Datenübertragung Im Ortsnetz," in *Ausgewählte Arbeiten über Nachrichtensysteme*, no. 56, Universität Erlangen-Nürnberg, 1984.
- [138] R. Rückriem, "Realisierung und messtechnische Untersuchung an einem digitalen Parallelverfahren zur Datenübertragung im Fernsprechkanal," in *Ausgewählte Arbeiten über Nachrichtensysteme*, no. 59, Universität Erlangen-Nürnberg, 1985.
- [139] J. Lindner et al, "OCDM – Ein Übertragungsverfahren für lokale Funknetze," in *Codierung fuer Quelle, Kanal und Uebertragung*, no. 130 in ITG Fachbericht, pp. pp 401–409, VDE Verlag, 26–28 Oct. 1994.
- [140] T. Keller, "Orthogonal frequency division multiplex techniques for wireless local area networks," 1996. Internal Report.
- [141] S. Nanda, D. J. Goodman, and U. Timor, "Performance of PRMA: A packet voice protocol for cellular systems," *IEEE Tr. on VT*, vol. 40, pp. 584–598, August 1991.
- [142] W. Webb, R. Steele, J. Cheung, and L. Hanzo, "A packet reservation multiple access assisted cordless telecommunications scheme," *IEEE Transactions on Veh. Technology*, vol. 43, pp. 234–245, May 1994.
- [143] R. A. 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.
- [144] M. Frullone, G. Riva, P. Grazioso, and C. Carciofy, "Investigation on dynamic channel allocation strategies suitable for PRMA schemes," *1993 IEEE Int. Symp. on Circuits and Systems, Chicago*, pp. 2216–2219, May 1993.

BIBLIOGRAPHY

871

- [145] M. Frullone, G. Falciasacca, P. Grazioso, G. Riva, and A. M. Serra, "On the performance of packet reservation multiple access with fixed and dynamic channel allocation," *IEEE Tr. on Veh. Techn.*, vol. 42, pp. 78–86, Feb. 1993.
- [146] J. Torrance, L. Hanzo, and T. Keller, "Interference resilience of burst-by-burst adaptive modems," in *Proceeding of ACTS Mobile Communication Summit '97* [482], pp. 489–494.
- [147] J. Torrance and L. Hanzo, "Statistical multiplexing for mitigating latency in adaptive modems," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'97* [483], pp. 938–942.
- [148] R. Hamming, "Error detecting and error correcting codes," *Bell Sys. Tech. J.*, 29, pp. 147–160, 1950.
- [149] P. Elias, "Coding for noisy channels," *IRE Conv. Rec. pt.4*, pp. 37–47, 1955.
- [150] J. Wozencraft, "Sequential decoding for reliable communication," *IRE Natl. Conv. Rec.*, vol. 5, pt.2, pp. 11–25, 1957.
- [151] J. Wozencraft and B. Reiffen, *Sequential decoding*. MIT Press, Cambridge, Mass., 1961.
- [152] R. Fano, "A heuristic discussion of probabilistic coding," *IEEE Trans. Info. Theory*, vol. IT-9, pp. 64–74, April 1963.
- [153] J. Massey, *Threshold decoding*. MIT Press, Cambridge, Mass., 1963.
- [154] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Info. Theory*, vol. IT-13, pp. 260–269, April 1967.
- [155] G. D. Forney, "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, pp. 268–278, March 1973.
- [156] J. Heller and I. Jacobs, "Viterbi decoding for satellite and space communication," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 835–848, October 1971.
- [157] 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*, pp. 1064–1070, May 1993.
- [158] A. Hocquenghem, "Codes correcteurs d'erreurs," *Chiffres (Paris)*, vol. 2, pp. 147–156, September 1959.
- [159] R. Bose and D. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Information and Control*, vol. 3, pp. 68–79, March 1960.
- [160] R. Bose and D. Ray-Chaudhuri, "Further results on error correcting binary group codes," *Information and Control*, vol. 3, pp. 279–290, September 1960.

- [161] W. Peterson, "Encoding and error correction procedures for the Bose-Chaudhuri codes," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 459–470, September 1960.
- [162] D. Gorenstein and N. Zierler, "A class of cyclic linear error-correcting codes in p^m symbols," *J. Soc. Ind. Appl. Math.*, 9, pp. 107–214, June 1961.
- [163] I. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. Soc. Ind. Appl. Math.*, vol. 8, pp. 300–304, June 1960.
- [164] E. Berlekamp, "On decoding binary Bose-Chaudhuri-Hocquenghem codes," *IEEE Trans. Info. Theory*, vol. 11, pp. 577–579, 1965.
- [165] E. Berlekamp, *Algebraic Coding Theory*. McGraw-Hill, New York, 1968.
- [166] J. Massey, "Step-by-step decoding of the Bose-Chaudhuri-Hocquenghem codes," *IEEE Trans. Info. Theory*, vol. 11, pp. 580–585, 1965.
- [167] J. Massey, "Shift-register synthesis and BCH decoding," *IEEE Tr. on Inf. Theory*, vol. IT-15, pp. 122–127, January 1969.
- [168] Consultative Committee for Space Data Systems, "Blue book," *Recommendations for Space Data System Standards: Telemetry Channel Coding*, May 1984.
- [169] W. Peterson, *Error correcting codes*. Cambridge, Mass, USA: MIT. Press, 1st ed., 1961.
- [170] W. Peterson and E. Weldon, Jr, *Error correcting codes*. MIT. Press, 2nd ed., August 1972. ISBN: 0262160390.
- [171] G. C. Clark, Jr and J. B. Cain, *Error correction coding for digital communications*. New York: Plenum Press, May 1981. ISBN: 0306406152.
- [172] A. Michelson and A. Levesque, *Error control techniques for digital communication*. J. Wiley and Sons, 1985.
- [173] R. Blahut, *Theory and practice of error control codes*. Addison-Wesley, 1983. ISBN 0-201-10102-5.
- [174] S. Lin and D. J. Costello Jr, *Error Control Coding: Fundamentals and Applications*. New Jersey, USA: Prentice-Hall, October 1982. ISBN: 013283796X.
- [175] V. Pless, *Introduction to the theory of error-correcting codes*. John Wiley and Sons, 1982. ISBN: 0471813044.
- [176] I. Blake, ed., *Algebraic coding theory: History and development*. Dowden, Hutchinson and Ross Inc., 1973.
- [177] K. Wong, *Transmission of channel coded speech and data over mobile channels*. PhD thesis, University of Southampton, 1989.
- [178] R. Steele, "Deploying personal communications networks," *IEEE Comms. Magazine*, pp. 12–15, September 1990.

BIBLIOGRAPHY

873

- [179] R. Lidl and H. Niederreiter, *Finite Fields*. Cambridge University Press, October 1996.
- [180] R. Steele, ed., *Mobile Radio Communications*. IEEE Press-Pentech Press, 1992.
- [181] D. Gorenstein and N. Zierler, "A class of error-correcting codes in p^m symbols," *J.Soc.Ind.Appl.Math.*, no. 9, pp. 207–214, 1961.
- [182] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, pp. 561–580, April 1975.
- [183] R. Blahut, *Fast algorithms for digital signal processing*. Addison-Wesley Publishing Company, 1985. ISBN 0-201-10155-6.
- [184] J. Schur, "Ueber Potenzreihen, die im Innern des Einheitskreises beschraenkt sind," *Journal fuer Mathematik*, pp. 205–232. Bd. 147, Heft 4.
- [185] R. Chien, "Cyclic decoding procedure for the Bose-Chaudhuri-Hocquenghem codes," *IEEE Trans. on Info. Theory*, vol. 10, pp. 357–363, October 1964.
- [186] A. Jennings, *Matrix computation for engineers and scientists*. J. Wiley and Sons Ltd., 1977.
- [187] G. Forney, Jr, "On decoding BCH codes," *IEEE Tr. on Inf. Theory*, vol. IT-11, pp. 549–557, 1965.
- [188] Y. Sugiyama, M. Kasahara, S. Hirasawa, and T. Namekawa, "A method for solving key equation for decoding goppa codes," *Inf. Control*, no. 27, pp. 87–99, 1975.
- [189] S. Golomb, *Shift register sequences*. Laugana Hills, CA: Aegean Park Press, 1982.
- [190] S. Lloyd, "Least squares quantisation in PCM," *Institute of Mathematical Statistics Meeting, Atlantic City, N.J.*, September 1957.
- [191] S. Lloyd, "Least squares quantisation in PCM," *IEEE Trans. on Information Theory*, vol. 28, no. 2, pp. 129–136, 1982.
- [192] J. Max, "Quantising for minimum distortion," *IRE Trans. on Information Theory*, vol. 6, pp. 7–12, 1960.
- [193] W. R. Bennett, "Spectra of quantised signals," *Bell System Technical Journal*, pp. 446–472, July 1946.
- [194] H. Holtzwarth, "Pulse code modulation und ihre verzerrung bei logarithmischer quanteilung," *Archiv der Elektrischen Uebertragung*, pp. 227–285, January 1949.
- [195] P. Panter and W. Dite, "Quantisation distortion in pulse code modulation with non-uniform spacing of levels," *Proc. of the IRE*, pp. 44–48, January 1951.
- [196] B. Smith, "Instantaneous companding of quantised signals," *Bell System Technical Journal*, pp. 653–709, 1957.

- [197] P. Noll and R. Zelinski, "A contribution to the quantisation of memoryless model sources," *Technical Report, Heinrich Heinz Institute, Berlin*, 1974. (in German).
- [198] M. Paez and T. Glisson, "Minimum mean squared error quantisation in speech PCM and DPCM systems," *IEEE Trans. on Communications*, pp. 225–230, April 1972.
- [199] A. K. Jain, *Fundamentals of Digital Image Processing*. Prentice-Hall, 1989.
- [200] R. A. Salami, *Robust Low Bit Rate Analysis-by-Synthesis Predictive Speech Coding*. PhD thesis, University of Southampton, 1990.
- [201] J. Makhoul, "Stable and efficient lattice methods for linear prediction," *IEEE Trans. on ASSP*, vol. 25, pp. 423–428, Oct. 1977.
- [202] N. Jayant, "Adaptive quantization with a one-word memory," *Bell System Technical Journal*, vol. 52, pp. 1119–1144, September 1973.
- [203] R. Steedman, "The common air interface MPT 1375," in Tuttlebee [489]. ISBN 3540196331.
- [204] L. Hanzo, "The British cordless telephone system: CT2," in Gibson [39], ch. 29, pp. 462–477.
- [205] H. Ochsner, "The digital european cordless telecommunications specification, DECT," in Tuttlebee [489], pp. 273–285. ISBN 3540196331.
- [206] S. Asghar, "Digital European Cordless Telephone," in Gibson [39], ch. 30, pp. 478–499.
- [207] "Personal handy phone (PHP) system." RCR Standard, STD-28, Japan.
- [208] "CCITT recommendation G.721."
- [209] 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.
- [210] A. H. Gray and J. D. Markel, "Distance measures for speech processing," *IEEE Transactions on ASSP*, vol. 24, no. 5, pp. 380–391, 1976.
- [211] 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, Feb. 1988.
- [212] P. Noll and R. Zelinski, "Bounds on quantizer performance in the low bit-rate region," *IEEE Transactions on Communications*, pp. 300–304, February 1978.
- [213] 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* [492], pp. 77–80.

BIBLIOGRAPHY

875

- [214] J. O'Neal, "Bounds on subjective performance measures for source encoding systems," *IEEE Transactions on Information Theory*, pp. 224–231, May 1971.
- [215] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proceedings of the IEEE*, pp. 1551–1588, November 1985.
- [216] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp. 247–254, June 1979.
- [217] J.-H. Chen, R. V. 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.
- [218] D. Sen and W. Holmes, "PERCELP-perceptually enhanced random codebook excited linear prediction," in *Proc. IEEE Workshop on Speech Coding for Telecommunications*, pp. 101–102, 1993.
- [219] 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* [491], pp. 1.3.1–1.3.4.
- [220] "Group speciale mobile (GSM) recommendation," April 1988.
- [221] S. Singhal and B. S. Atal, "Amplitude optimization and pitch prediction in multipulse coders," *IEEE Trans. on Acoustics, Speech and Signal Processing*, pp. 317–327, Mar 1989.
- [222] "Federal standard 1016 – telecommunications: Analog to digital conversion of radio voice by 4,800 bits/second code excited linear prediction (CELP)," February 14 1991.
- [223] S. Wang and A. Gersho, "Phonetic segmentation for low rate speech coding," in Atal *et al.* [28], pp. 257–266. ISBN: 0792390911.
- [224] 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)* [496], pp. 143–146.
- [225] D. W. Griffin and J. S. Lim, "Multiband excitation vocoder," *IEEE Trans. on Acoustics, Speech and Signal Processing*, pp. 1223–1235, August 1988.
- [226] 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)* [496], pp. 151–154.
- [227] W. B. Kleijn, "Encoding speech using prototype waveforms," *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 386–399, October 1993.

- [228] V. Ramamoorthy and N. Jayant, "Enhancement of ADPCM speech by adaptive postfiltering," *Bell Syst Tech Journal*, vol. 63, pp. 1465–1475, October 1984.
- [229] 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.
- [230] 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* [494], pp. 2185–2188.
- [231] ITU-T, *CCITT Recommendation G.728: Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction*, 1992.
- [232] 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.
- [233] F. Itakura and S. Saito, "Analysis-synthesis telephony based upon the maximum likelihood method," in *Proc. of the 6th International Congress on Acoustic*, (Tokyo), pp. C17–20, 1968.
- [234] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electr. and Comms. in Japan*, vol. 53-A, pp. 36–43, 1970.
- [235] N. Kitawaki, K. Itoh, and F. Itakura, "PARCOR speech analysis synthesis system," *Review of the Electr. Comm. Lab., Nippon TTPC*, vol. 26, pp. 1439–1455, Nov-Dec 1978.
- [236] R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE Trans. on ASSP*, pp. 309–321, 1975.
- [237] 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.
- [238] 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, January 1993.
- [239] F. K. 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* [491], pp. 1.10.1–1.10.4.
- [240] G. Kang and L. Fransen, "Low-bit rate speech encoders based on line-spectrum frequencies (LSFs)," Tech. Rep. 8857, NRL, November 1984.
- [241] P. Kabal and R. Ramachandran, "The computation of line spectral frequencies using chebyshev polynomials," *IEEE Trans. ASSP*, vol. 34, pp. 1419–1426, December 1986.

BIBLIOGRAPHY

877

- [242] M. Omologo, "The computation and some spectral considerations on line spectrum pairs (LSP)," in *Proc. EUROSPEECH*, pp. 352–355, 1989.
- [243] B. Cheetham, "Adaptive LSP filter," *Electronics Letters*, vol. 23, pp. 89–90, January 1987.
- [244] K. Geher, *Linear Circuits*. Budapest, Hungary: Technical Publishers, 1972. (in Hungarian).
- [245] 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.
- [246] A. Lepschy, G. Mian, and U. Viaro, "A note on line spectral frequencies," *IEEE Trans. ASSP*, vol. 36, pp. 1355–1357, August 1988.
- [247] B. Cheetham and P. Huges, "Formant estimation from LSP coefficients," in *Proc. IERE 5th Int. Conf. on Digital Processing of Signals in Communications*, pp. 183–189, 20–23 Sept 1988.
- [248] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.
- [249] 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* [494], pp. 2181–2184.
- [250] R. Ramachandran, M. Sondhi, N. Seshadri, and B. Atal, "A two codebook format for robust quantisation of line spectral frequencies," *IEEE Trans. on Speech and Audio Processing*, vol. 3, pp. 157–168, May 1995.
- [251] 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)* [496], pp. 1–4.
- [252] K. Lee, A. Kondo, and B. Evans, "Speaker adaptive vector quantisation of LPC parameters of speech," *Electronic Letters*, vol. 24, pp. 1392–1393, October 1988.
- [253] 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* [494], pp. 2404–2407.
- [254] R. A. 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* [492], pp. 124–128.
- [255] 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* [495], pp. 402–405.

- [256] J. Huang and P. Schultheis, "Block quantization of correlated gaussian random variables," *IEEE Trans. Commun. Sys.*, vol. 11, pp. 289–296, September 1963.
- [257] R. A. Salami, L. Hanzo, and D. Appleby, "A computationally efficient CELP codec with stochastic vector quantization of LPC parameters," in *URSI Int. Symposium on Signals, Systems and Electronics*, (Erlangen, West Germany), pp. 140–143, 18–20 Sept 1989.
- [258] 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* [492], pp. 69–72.
- [259] 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* [493], pp. 641–644.
- [260] H. Harborg, J. Knudson, A. Fudseth, and F. Johansen, "A real time wideband CELP coder for a videophone application," in *Proceedings of ICASSP*, pp. II121 – II124, 1994.
- [261] R. Lefebvre, R. Salami, C. Laflamme, and J. Adoul, "High quality coding of wideband audio signals using transform coded excitation (TCX)," in *Proceedings of ICASSP*, pp. I193–I196, 1994.
- [262] J. Paulus and J. Schnitzler, "16kbit/s wideband speech coding based on unequal subbands," in *Proceedings of ICASSP*, pp. 255–258, 1996.
- [263] J. Chen and D. Wang, "Transform predictive coding of wideband speech signals," in *Proceedings of ICASSP*, pp. 275–278, 1996.
- [264] A. Ubale and A. Gersho, "A multi-band CELP wideband speech coder," in *Proceedings of ICASSP*, pp. 1367–1370, 1997.
- [265] 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 ICASSP*, 1999.
- [266] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *Journal of the Acoustic Society of America*, vol. 57, p. S35, 1975.
- [267] 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.
- [268] "7 khz audio coding within 64 kbit/s." CCITT Recommendation G.722, 1988.
- [269] "Recommendation G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)." CCITT Study Group XVIII, June 30 1995. Version 6.31.

BIBLIOGRAPHY

879

- [270] T. Eriksson, J. Linden, and J. Skoglung, "A safety-net approach for improved exploitation of speech correlation," in *Proceedings of ICASSP*, pp. 96–101, 1995.
- [271] T. Eriksson, J. Linden, and J. Skoglung, "Exploiting interframe correlation in spectral quantization - a study of different memory VQ schemes," in *Proceedings of ICASSP*, pp. 765–768, May 1996.
- [272] H. Zarrinkoub and P. Mermelstein, "Switched prediction and quantization of LSP frequencies," in *Proceedings of ICASSP*, pp. 757–764, May 1996.
- [273] J. E. 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.
- [274] J. Schur, "Über potenzreihen, die im innern des einheitskreises beschränkt sind," *Journal für die reine und angewandte Mathematik*, Bd 14, pp. 205–232, 1917.
- [275] W. Webb, L. Hanzo, R. A. 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)* [490], pp. 511–516.
- [276] L. Hanzo, W. Webb, R. A. Salami, and R. Steele, "On QAM speech transmission schemes for microcellular mobile PCNs," *European Transactions on Communications*, pp. 495–510, Sept/Oct 1993.
- [277] J. Williams, L. Hanzo, R. Steele, and J. Cheung, "A comparative study of microcellular speech transmission schemes," *IEEE Tr. on Veh. Technology*, vol. 43, pp. 909–925, Nov 1994.
- [278] "Cellular system dual-mode mobile station-base station compatibility standard IS-54B." Telecommunications Industry Association Washington DC, 1992. EIA/TIA Interim Standard.
- [279] A. Black, A. Kondo, and B. Evans, "High quality low delay wideband speech coding at 16 kbit/sec," in *Proc of 2nd Int Workshop on Mobile Multimedia Communications*, 11-14 April 1995. Bristol University, UK.
- [280] C. Laflamme, J.-P. Adoul, R. A. Salami, S. Morissette, and P. Mabillean, "16 Kbps wideband speech coding technique based on algebraic CELP," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'91* [493], pp. 13–16.
- [281] R. A. Salami, C. Laflamme, and J.-P. Adoul, "Real-time implementation of a 9.6 kbit/s ACELP wideband speech coder," in *Proc GLOBECOM '92*, 1992.
- [282] I. Gerson and M. Jasiuk, "Vector sum excited linear prediction (VSELP)," in Atal *et al.* [28], pp. 69–80. ISBN: 0792390911.
- [283] 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* [492], pp. 57–60.

- [284] 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* [492], pp. 61–64.
- [285] 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 *Proc. of IERE 5th Int. Conf. on Digital Processing of Signals in Comms*, pp. 167–174, September 1988.
- [286] J. Adoul, P. Mabillean, 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* [494], pp. 1957–1960.
- [287] 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, Nov 1995.
- [288] H. Law and R. Seymour, "A reference distortion system using modulated noise," *IEE Paper*, pp. 484–485, Nov. 1962.
- [289] 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* [495], pp. 147–150.
- [290] Y. Tohkura, F. Itakura, and S. Hashimoto, "Spectral smoothing technique in PARCOR speech analysis-synthesis," *IEEE Trans. on Acoustics, Speech and Signal Processing*, pp. 587–596, 1978.
- [291] J.-H. Chen and R. V. Cox, "Convergence and numerical stability of backward-adaptive LPC predictor," in *Proceedings of IEEE Workshop on Speech Coding for Telecommunications*, pp. 83–84, 1993.
- [292] S. Singhal and B. S. Atal, "Optimizing LPC filter parameters for multi-pulse excitation," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'83* [497], pp. 781–784.
- [293] 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* [498], pp. 73–76.
- [294] G. H. Golub and C. F. V. Loan, "An analysis of the total least squares problem," *SIAM Journal of Numerical Analysis*, vol. 17, no. 6, pp. 883–890, 1980.
- [295] 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.
- [296] R. D. Degroat and E. M. Dowling, "The data least squares problem and channel equalization," *IEEE Transactions on Signal Processing*, pp. 407–411, 1993.

BIBLIOGRAPHY

881

- [297] F. Tzeng, "Near-optimum linear predictive speech coding," in *IEEE Global Telecommunications Conference*, pp. 508.1.1–508.1.5, 1990.
- [298] M. Niranjan, "CELP coding with adaptive output-error model identification," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'90* [485], pp. 225–228.
- [299] 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)* [486], pp. 114–118.
- [300] R. V. Cox, W. B. 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* [492], pp. 739–742.
- [301] J. P. 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* [492], pp. 735–738.
- [302] S. Atungsiri, A. Kondo, and B. Evans, "Error control for low-bit-rate speech communication systems," *IEE Proceedings-I*, vol. 140, pp. 97–104, April 1993.
- [303] L. Ong, A. Kondo, and B. Evans, "Enhanced channel coding using source criteria in speech coders," *IEE Proceedings-I*, vol. 141, pp. 191–196, June 1994.
- [304] W. Kleijn, "Source-dependent channel coding and its application to CELP," in Atal *et al.* [28], pp. 257–266. ISBN: 0792390911.
- [305] J. Woodard and L. Hanzo, "A dual-rate algebraic CELP-based speech transceiver," in *Proceedings of IEEE VTC '94* [480], pp. 1690–1694.
- [306] 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* [485], pp. 177–180.
- [307] J. Williams, L. Hanzo, and R. Steele, "Channel-adaptive voice communications," in *Proceedings of IEE Conference on Radio Receivers and Associated Systems (RRAS'95)* [486], pp. 144–147.
- [308] T. E. Tremain, "The government standard linear predictive coding algorithm: LPC-10," *Speech Technology*, vol. 1, pp. 40–49, April 1982.
- [309] J. P. Campbell, T. E. Tremain, and V. C. Welch, "The DoD 4.8 kbps standard (proposed federal standard 1016)," in Atal *et al.* [28], pp. 121–133. ISBN: 0792390911.
- [310] 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* [485], pp. 665–668.

- [311] 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, Oct 1988.
- [312] Y. Shoham, "Constrained-stochastic excitation coding of speech at 4.8 kb/s," in Atal *et al.* [28], pp. 339–348. ISBN: 0792390911.
- [313] 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.
- [314] 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* [485], pp. 461–464.
- [315] I. Gerson and M. Jasiuk, "Techniques for improving the performance of CELP-type speech codecs," *IEEE JSAC*, vol. 10, pp. 858–865, June 1992.
- [316] I. Gerson, "Method and means of determining coefficients for linear predictive coding." US Patent No 544,919, October 1985.
- [317] A. Cumain, "On a covariance-lattice algorithm for linear prediction," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'82* [499], pp. 651–654.
- [318] 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. S. Atal, V. Cuperman, and A. Gersho, eds.), pp. 85–92, Kluwer Academic Publishers, 1993.
- [319] K. Mano, T. Moriya, S. Miki, H. Ohmuro, K. Ikeda, and J. Ikeda, "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.
- [320] 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.
- [321] 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)* [500], pp. 93–96.
- [322] R. A. 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)* [500], pp. 97–100.
- [323] 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* [482], pp. 894–899.

BIBLIOGRAPHY

883

- [324] T. Ojanpare et al, "FRAMES multiple access technology," in *Proceedings of IEEE ISSSTA '96*, vol. 1, (Mainz, Germany), pp. 334–338, IEEE, Sept 1996.
- [325] 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.
- [326] 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.
- [327] P. Jung and M. NaBhan, "Performance evaluation of turbo codes for short frame transmission systems," *IEE Electronic Letters*, pp. 111–112, Jan 1994.
- [328] A. Barbulescu and S. Pietrobon, "Interleaver design for turbo codes," *IEE Electronic Letters*, pp. 2107–2108, Dec 1994.
- [329] 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.
- [330] "COST 207: Digital land mobile radio communications, final report." Office for Official Publications of the European Communities, 1989. Luxembourg.
- [331] R. A. 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)* [501], pp. 775–778.
- [332] R. A. 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, Sept 1997.
- [333] R. A. Salami, C. Laflamme, B. Besette, J.-P. Adoul, K. Jarvinen, J. Vainio, P. Kapanen, T. Hakanen, and P. Haavisto, "Description of the GSM enhanced full rate speech codec," in *Proc. of ICC'97*, 1997.
- [334] "PCS1900 enhanced full rate codec US1." SP-3612.
- [335] "IS-136.1A TDMA cellular/PCS - radio interface - mobile station - base station compatibility digital control channel." Revision A, Aug. 1996.
- [336] T. Honkanen, J. Vainio, K. Jarvinen, P. Haavisto, R. A. 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)* [501], pp. 731–734.
- [337] "TIA/EIA/IS641, interim standard, TDMA cellular/PCS radio intergface - enhanced full-rate speech codec," May 1996.

- [338] "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s." CCITT Recommendation G.723.1, March 1996.
- [339] C. Hong, *Low Delay Switched Hybrid Vector Excited Linear Predictive Coding of Speech*. PhD thesis, National University of Singapore, 1994.
- [340] 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.
- [341] J.-H. Chen, N. Jayant, and R. V. 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* [498].
- [342] 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.
- [343] 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.
- [344] J.-H. Chen, Y.-C. Lin, and R. V. Cox, "A fixed-point 16 kb/s LD-CELP algorithm," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'91* [493], pp. 21–24.
- [345] 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* [485], pp. 453–456.
- [346] 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.
- [347] K. Zeger and A. Gersho, "Zero-redundancy channel coding in vector quantization," *Electr. Letters*, vol. 23, pp. 654–656, June 1987.
- [348] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantiser design," *IEEE Transactions on Communications*, vol. Com-28, January 1980.
- [349] W. B. Kleijn, D. J. Krasinski, and R. H. Ketchum, "Fast methods for the CELP speech coding algorithm," *IEEE Trans. on Acoustics, Speech and Signal Processing*, pp. 1330–1342, August 1990.
- [350] S. L. Dall'Agnol, J. R. B. 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* [480], pp. 1699–1703.
- [351] X. Maitre, "7 khz audio coding within 64 kbit/s," *IEEE-JSAC*, vol. 6, pp. 283–298, February 1988.

BIBLIOGRAPHY

885

- [352] R. Crochiere, S. Webber, and J. Flanagan, "Digital coding of speech in sub-bands," *Bell System Tech. Journal*, pp. 1069–1085, October 1976.
- [353] R. Crochiere, "An analysis of 16 kbit/s sub-band coder performance: dynamic range, tandem connections and channel errors," *BSTJ*, vol. 57, pp. 2927–2952, October 1978.
- [354] 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*, (Hartford, Conn, USA), pp. 191–195, IEEE, 9–11 May 1977.
- [355] 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* [488], pp. 291–294.
- [356] H. Nussbaumer, "Complex quadrature mirror filters," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'83* [497], pp. 221–223.
- [357] C. Galand and H. Nussbaumer, "New quadrature mirror filter structures," *IEEE Trans. on ASSP*, vol. ASSP-32, pp. 522–531, June 1984.
- [358] 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* [493], pp. 1–4.
- [359] J. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE-JSAC*, vol. 6, no. 2, pp. 314–323, 1988.
- [360] 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* [493], pp. 9–12.
- [361] R. Soheili, A. Kondo, 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.
- [362] V. Sanchez-Calle, C. Laflamme, R. A. 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, Elsevier Science Publishers, 1992.
- [363] 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* [493], pp. 17–20.
- [364] L. Hanzo, W. Webb, and T. Keller, *Single- and Multi-carrier Quadrature Amplitude Modulation*. IEEE Press-Pentech Press, April 2000.

- [365] 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* [483], pp. 472–477.
- [366] C. H. Wong, T. H. Liew, and L. Hanzo, "Turbo coded burst by burst adaptive wideband modulation with blind modem mode detection," in Electronic copy [478], pp. 303–308.
- [367] M. S. Yee, T. H. Liew, and L. Hanzo, "Radial basis function decision feedback equalisation assisted block turbo burst-by-burst adaptive modems," in *Proceeding of VTC'99 (Fall)*, (Amsterdam, Netherlands), pp. 1600–1604, IEEE, 19–22 September 1999.
- [368] 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* [484], pp. 487–491.
- [369] V. K. N. Lau and M. D. 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)* [481], pp. 348–352.
- [370] T. Keller and L. Hanzo, "Adaptive orthogonal frequency division multiplexing schemes," in *Proceeding of ACTS Mobile Communication Summit '98*, (Rhodes, Greece), pp. 794–799, ACTS, 8–11 June 1998.
- [371] E. L. Kuan, C. H. Wong, and L. Hanzo, "Burst-by-burst adaptive joint detection CDMA," in *Proceeding of VTC'99 (Spring)* [479].
- [372] K. Fazel and G. Fettweis, eds., *Multi-carrier spread-spectrum*. Kluwer, 1997. p260, ISBN 0-7923-9973-0.
- [373] T. May and H. Rohling, "Reduktion von Nachbarkanalstörungen in OFDM-Funkübertragungssystemen," in *2. OFDM-Fachgespräch in Braunschweig*, 1997.
- [374] S. H. Müller and J. B. Huber, "Vergleich von OFDM-Verfahren mit reduzierter Spitzenleistung," in *2. OFDM-Fachgespräch in Braunschweig*, 1997.
- [375] 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.
- [376] 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* [480], pp. 1655–1659.
- [377] S. J. Shepherd, P. W. J. van Eetvelt, C. W. Wyatt-Millington, and S. K. Barton, "Simple coding scheme to reduce peak factor in QPSK multicarrier modulation," *Electronics Letters*, vol. 31, pp. 1131–1132, July 1995.

BIBLIOGRAPHY

887

- [378] A. E. Jones, T. A. Wilkinson, and S. K. 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.
- [379] M. D. Benedetto and P. Mandarini, "An application of MMSE predistortion to OFDM systems," *IEEE Trans. on Comm.*, vol. 44, pp. 1417–1420, Nov 1996.
- [380] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "A practical discrete multi-tone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. on Communications*, vol. 48, pp. 772–775, 1995.
- [381] K. Fazel, S. Kaiser, P. Robertson, and M. J. Ruf, "A concept of digital terrestrial television broadcasting," *Wireless Personal Communications*, vol. 2, pp. 9–27, 1995.
- [382] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial tv broadcasting," *IEEE Communications Magazine*, pp. 100–109, February 1995.
- [383] 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* [482], pp. 653–658.
- [384] Y. Li and N. R. Sollenberger, "Interference suppression in OFDM systems using adaptive antenna arrays," in *Proceeding of Globecom'98*, (Sydney, Australia), pp. 213–218, IEEE, 8–12 Nov 1998.
- [385] F. W. Vook and K. L. Baum, "Adaptive antennas for OFDM," in *Proceedings of IEEE Vehicular Technology Conference (VTC'98)* [481], pp. 608–610.
- [386] T. Keller, J. Woodard, and L. Hanzo, "Turbo-coded parallel modem techniques for personal communications," in *Proceedings of IEEE VTC '97* [487], pp. 2158–2162.
- [387] T. Keller and L. Hanzo, "Blind-detection assisted sub-band adaptive turbo-coded OFDM schemes," in *Proceeding of VTC'99 (Spring)* [479], pp. 489–493.
- [388] "Universal mobile telecommunications system (UMTS); UMTS terrestrial radio access (UTRA); concept evaluation," tech. rep., ETSI, 1997. TR 101 146.
- [389] tech. rep. <http://standards.pictel.com/ptelcont.htm#Audio>.
- [390] M. Failli, "Digital land mobile radio communications COST 207," tech. rep., European Commission, 1989.
- [391] H. S. Malvar, *Signal Processing with Lapped Transforms*. Artech House, Boston, MA, 1992.
- [392] K. Rao and P. Yip, *Discrete cosine transform: algorithms, advantages and applications*. Academic Press Ltd., UK, 1990.

- [393] B. Atal and M. Schroeder, "Predictive coding of speech signals," *Bell System Technical Journal*, pp. 1973–1986, October 1970.
- [394] 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.
- [395] 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.
- [396] 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)* [503], pp. 488–491.
- [397] 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.
- [398] 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.
- [399] 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)* [496], pp. 439–442.
- [400] 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.
- [401] 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)* [503], pp. 500–503.
- [402] 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)* [503], pp. 508–511.
- [403] 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* [485], pp. 641–643.
- [404] K. Teague, W. Andrews, and B. Walls, "Harmonic speech coding at 2400 bps," in *Proc. 10th Annual Mid-America Symposium on Emerging Computer Technology*, (Norman, Oklahoma, USA), 1996.

BIBLIOGRAPHY

889

- [405] 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.
- [406] 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)* [502], pp. 200–203.
- [407] 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* [498], pp. 137–140.
- [408] 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.
- [409] 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)* [502], pp. 212–215.
- [410] 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.
- [411] 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.
- [412] 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.
- [413] W. Voiers, "Diagnostic acceptability measure for speech communication systems," in *Proceedings of ICASSP 77*, pp. 204–207, May 1977.
- [414] W. Voiers, "Evaluating processed speech using the diagnostic rhyme test," *Speech Technology*, January/February 1983.
- [415] 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)* [502], pp. 1137–1140.
- [416] 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)* [502], pp. 1141–1144.
- [417] J. Tardelli and E. W. Kreamer, "Vocoder intelligibility and quality test methods," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [502], pp. 1145–1148.

- [418] 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)* [502], pp. 1149–1152.
- [419] E. W. Creamer and J. Tardelli, "Communicability testing for voice coders," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)* [502], pp. 1153–1156.
- [420] B. Atal and L. Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, pp. 201–212, June 1976.
- [421] 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.
- [422] A. Noll, "Cepstrum pitch determination," *Journal of the Acoustical Society of America*, vol. 41, pp. 293–309, February 1967.
- [423] 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.
- [424] L. Rabiner, M. Cheng, A. Rosenberg, and C. McGonegal, "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.
- [425] DVSI, *Inmarsat-M Voice Codec*, Issue 3.0 ed., August 1991.
- [426] 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.
- [427] 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.
- [428] T. Koornwinder, *Wavelets: An Elementary Treatment of Theory and Applications*. World Scientific, 1993.
- [429] C. Chui, *Wavelet Analysis and its Applications*, vol. I: An Introduction to Wavelets. Academic Press, 1992.
- [430] C. Chui, *Wavelet Analysis and its Applications*, vol. II: Wavelets: A Tutorial in Theory and Applications. Academic Press, 1992.
- [431] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Magazine*, pp. 14–38, October 1991.
- [432] A. Graps, "An introduction to wavelets," *IEEE Computational Science & Engineering*, pp. 50–61, Summer 1995.

BIBLIOGRAPHY

891

- [433] A. Cohen and J. K. Cević, "Wavelets: The mathematical background," *Proceedings of the IEEE*, vol. 84, pp. 514–522, April 1996.
- [434] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, pp. 961–1005, September 1990.
- [435] 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.
- [436] H. Baher, *Analog & Digital Signal Processing*. John Wiley & Sons, 1990.
- [437] 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)* [502], pp. 546–549.
- [438] 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.
- [439] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proceedings of the IEEE*, vol. 84, pp. 626–638, April 1996.
- [440] S. Mallat and W. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, pp. 617–643, March 1992.
- [441] M. Vetterli and J. K. Cević, *Wavelets and Subband Coding*. Prentice-Hall, 1995.
- [442] 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.
- [443] F. Brooks, B. Yeap, J. Woodard, and L. Hanzo, "A sixth-rate, 3.8kbps gsm-like speech transceiver," in *Proceedings of ACTS summit*, pp. 647–652, 1998.
- [444] F. Brooks, E. Kuan, and L. Hanzo, "A 2.35kbps joint-detection CDMA speech transceiver," in *Proceedings of VTC '99*, pp. 2403–2407, 1999.
- [445] 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.
- [446] P. Robertson, "Illuminating the structure of code and decoder of parallel concatenated recursive systematic (turbo) codes," *IEEE Globecom*, pp. 1298–1303, 1994.
- [447] 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, Dec 1990.

- [448] J. Erfanian, S. Pasupathy, and G. Gulak, "Reduced complexity symbol detectors with parallel structures for ISI channels," *IEEE Transactions on Communications*, vol. 42, pp. 1661–1671, 1994.
- [449] J. Hagenauer and P. Hoeher, "A viterbi algorithm with soft-decision outputs and its applications," in *IEEE Globecom*, pp. 1680–1686, 1989.
- [450] 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.
- [451] L. Rabiner, C. McGonegal, and D. Paul, *FIR Windowed Filter Design Program - WINDOW*, ch. 5.2. IEEE Press, 1979.
- [452] S. Yeldner, A. Kondo, and B. Evans, "Multiband linear predictive speech coding at very low bit rates," *IEEE Proceedings in Vision, Image and Signal Processing*, vol. 141, pp. 284–296, October 1994.
- [453] 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* [483], pp. 37–41.
- [454] J. Flanagan and R. Golden, "Phase vocoder," *The Bell System Technical Journal*, pp. 1493–1509, November 1966.
- [455] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on sinusoidal representation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, pp. 744–754, August 1986.
- [456] 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.
- [457] 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.
- [458] 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.
- [459] 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.
- [460] R. McAulay and T. Quatieri, "Sinusoidal coding," in *Speech Coding and Synthesis* (W.B.Keijn and K.K.Paliwal, eds.), ch. 4, Elsevier Science, 1995.
- [461] 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, Kluwer Academic Publishers, 1991.

BIBLIOGRAPHY

893

- [462] M. Nishiguchi and J. Matsumoto, "Harmonic and noise coding of LPC residuals with classified vector quantization," in *Proceedings of ICASSP 95*, pp. 484–487, 1995.
- [463] V. Cuperman, P. Lupini, and B. Bhattacharya, "Spectral excitation coding of speech at 2.4kb/s," in *Proceedings of ICASSP95*, pp. 496–499, 1995.
- [464] S. Yeldner, A. Kondo, and B. Evans, "High quality multiband LPC coding of speech at 2.4kbit/s," *Electronics Letters*, vol. 27, no. 14, pp. 1287–1289, 1991.
- [465] 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)* [503], pp. 516–518.
- [466] 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)* [501], pp. 1579–1582.
- [467] 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)* [501], pp. 1555–1558.
- [468] 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.
- [469] K. Kryter, "Methods for the calculation of the articulation index," tech. rep., American National Standards Institute, 1965.
- [470] U. Halka and U. Heute, "A new approach to objective quality-measures based on attribute matching," *Speech Communications*, vol. 11, pp. 15–30, 1992.
- [471] 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.
- [472] 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.
- [473] 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* [488], pp. 706–709.
- [474] P. Breitkopf and T. Barnwell III, "Segmental preclassification for improved objective speech quality measures," in *IEEE Proc. of Internal. Conf. Acoust. Speech Signal Process.*, pp. 1101–1104, 1981.

- [475] L. Hanzo and L. Hinsenkamp, "On the subjective and objective evaluation of speech codecs," *Budavox Telecommunications Review*, no. 2, pp. 6–9, 1987.
- [476] K. D. Kryter, "Masking and speech communications in noise," in *The effects of Noise on Man*, ch. 2, Academic Press, 1970. ISBN: 9994669966.
- [477] A. House, C. Williams, M. Hecker, and K. Kryter, "Articulation testing methods: Consonated differentiation with a closed-response set," *J Acoust. Soc. Am.*, pp. 158–166, Jan. 1965.
- [478] ACTS, *Proceeding of ACTS Mobile Communication Summit '99*, (Sorrento, Italy), June 8–11 1999.
- [479] IEEE, *Proceeding of VTC'99 (Spring)*, (Houston, Texas, USA), 16–20 May 1999.
- [480] IEEE, *Proceedings of IEEE VTC '94*, (Stockholm, Sweden), June 8–10 1994.
- [481] IEEE, *Proceedings of IEEE Vehicular Technology Conference (VTC'98)*, (Ottawa, Canada), May 1998.
- [482] ACTS, *Proceeding of ACTS Mobile Communication Summit '97*, (Aalborg, Denmark), 7–10 October 1997.
- [483] IEEE, *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'97*, (Marina Congress Centre, Helsinki, Finland), 1–4 Sept 1997.
- [484] IEEE, *Proceedings of IEEE VTC '96*, (Atlanta, GA, USA), 1996.
- [485] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'90*, (Albuquerque, New Mexico, USA), 3–6 April 1990.
- [486] IEE, *Proceedings of IEE Conference on Radio Receivers and Associated Systems (RRAS'95)*, (Bath, UK), 26–28 September 1995.
- [487] IEEE, *Proceedings of IEEE VTC '97*, (Phoenix, Arizona, USA), 4–7 May 1997.
- [488] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'80*, (Denver, Colorado, USA), 9–11 April 1980.
- [489] W. H. Tuttlebee, ed., *Cordless telecommunications in Europe : the evolution of personal communications*. London: Springer-Verlag, 1990. ISBN 3540196331.
- [490] IEEE, *Proceedings of IEEE Vehicular Technology Conference (VTC'91)*, (St. Louis, MO, USA), 19–22 May 1991.
- [491] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'84*, (San Diego, California, USA), 19–21 March 1984.
- [492] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'89*, (Glasgow, Scotland, UK), 23–26 May 1989.

BIBLIOGRAPHY

895

- [493] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'91*, (Toronto, Ontario, Canada), 14–17 May 1991.
- [494] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'87*, (Dallas, TX, USA), 6–9 April 1987.
- [495] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'88*, (New York, NY, USA), 11–14 April 1988.
- [496] IEEE, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'93)*, (Minneapolis, MN, USA), 27–30 Apr 1993.
- [497] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'83*, (Boston, Mass, USA), 14–16 April 1983.
- [498] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'92*, March 1992.
- [499] IEEE, *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, ICASSP'82*, May 1982.
- [500] IEEE, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'94)*, (Adelaide, Australia), 19–22 Apr 1994.
- [501] IEEE, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'97)*, (Munich, Germany), 21–24 April 1997.
- [502] IEEE, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)*, (Atlanta, USA), May 7-10 1996.
- [503] IEEE, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'95)*, (Detroit, MI, USA), 9–12 May 1995.

Author Index

A

F. Adachi [84].....59, 68
Patrick Adde [450].....730, 762
J-P. Adoul [287].....336
J-P. Adoul [306].....370
J-P. Adoul [280]..322, 438, 530, 575–578,
581
J-P. Adoul [333].....454, 458, 459, 463
J-P. Adoul [281]..322, 576, 579–581, 599
J-P. Adoul [322].....428, 518
J-P. Adoul [331].....451, 453
J-P. Adoul [362].....577, 578, 581
J.P. Adoul [286]..328, 334, 335, 454, 459,
460, 463, 576
J.P. Adoul [284].....328
J.P. Adoul [261].....282
J.P. Adoul [336].....460, 463
Jean-Pierre Adoul [143] 80, 321, 336, 428,
463, 481, 518, 526
Jean-Pierre Adoul [332].....451, 453
A. H. Aghvami [117].....75, 76
M. Alard [132].....76
Abraham Alcaim [350].....536
A. Aldroubi [439].....666
L.B. Almeida [456].....769
L.B. Almeida [310].....394
M-S. Alouini [107].....69, 580
W. Andrews [397].....607, 610
W. Andrews [404].....610
Ettiboua Angui [450].....730, 762
D.G. Appleby [254].....273, 275
D.G. Appleby [257].....279, 327
K. Arimochi [365].....580
Saf Asghar [206].....204
B.S. Atal [395].....601–603, 733
B.S. Atal [420].....634, 666
B.S. Atal [253].....273, 275

B.S. Atal [258].....279
B.S. Atal [250].....273, 279, 280
B.S. Atal [412].....619
B.S. Atal [219].....240
Bishnu S. Atal [216].....237, 238
Bishnu S. Atal [9].....244, 301, 603
Bishnu S. Atal [28].....601
Bishnu S. Atal [238]..261, 273, 278, 279,
286, 357, 629
Bishnu S. Atal [16]245, 321–323, 501, 604
Bishnu S. Atal [292].....349
Bishnu S. Atal [221].....244
S.A. Atungsiri [302].....358, 359

B

D.K. Baghbadrani [285].....328, 576
H. Baher [436].....663
L.R. Bahl [329].....445, 585, 730
A. Baier [447].....730
A.S. Barbulescu [328].....445
T.P. Barnwell [406].....612
S. K. Barton [378].....582
S. K. Barton [377].....582
A. Bateman [91].....64
A. Bateman [86].....59, 64
A. Bateman [85].....59
K. L. Baum [385].....582
M.G. Di Benedetto [379].....582
W. R. Bennett [193].....172
E.R. Berlekamp [164].....85, 86
E.R. Berlekamp [165]...86, 99, 100, 118,
130, 141
Claude Berrou [450].....730, 762
Claude Berrou [157]...85, 444, 582, 584,
726–728, 762
Claude Berrou [325].....444, 582, 584,
726–728, 762

AUTHOR INDEX

897

- B. Besette [333] 454, 458, 459, 463
 B. Bessette [331] 451, 453
 Bruno Bessette [332] 451, 453
 B. Bhattacharya [463] 776
 M.R. Bielefeld [416] 622
 J. A. C. Bingham [380] 582
 A.W. Black [279] 322, 572, 573, 575, 580,
 581, 598
 R.E. Blahut [173] . 99, 100, 118, 120, 121,
 123, 130, 131, 134, 141, 153
 R.E. Blahut [183] 123, 153
 I.F. Blake [176] 99, 120, 130
 S.E. Blumstein [20] 602
 H. Bochmann [129] 76
 E. Bogenfeld [383] 582
 J. Borowski [383] 582
 R.C. Bose [159] 85
 R.C. Bose [160] 85
 G.F. Boudreaux-Bartels [423]... 635, 665,
 666, 671, 675, 676
 A. Brind'Amour [398] 607, 608, 776
 D.P. Brock [418] 622
 F.C.A. Brooks [443] 719
 F.C.A. Brooks [444] 719
 K. Bryden [398] 607, 608, 776
 A. G. Burr [98] 68
 A. Buzo [348] 509, 574
- C**
- J. Bibb Cain [171] 99, 118, 130, 131, 141,
 146
 Joseph P. Campbell [301] . 357, 358, 362,
 392
 Joseph P. Campbell [309] . 392, 396, 637
 C. Carciofy [144] 81, 380
 K.W. Cattermole [4] 175, 176, 601
 J.K. Cavers [92] 64
 J.K. Cavers [83] 59–61, 64, 68
 J.K. Cavers [73] 43
 J. Kovačević [433] 663
 J. Kovačević [441] 666
 T. Champion [403] 609, 776, 779
 T.G. Champion [415] 622
 R.W. Chang [120] 76, 582
 R.W. Chang [124] 76
 B.M.G. Cheetham [243] 265
 B.M.G. Cheetham [247] 269
 J.H. Chen [263] 282
 Juin-Hwey Chen [344] 487
 Juin-Hwey Chen [341] 482
 Juin-Hwey Chen [343] 482, 484
 Juin-Hwey Chen [230] 246, 397, 482, 485,
 497
 Juin-Hwey Chen [342] 482, 509
 Juin-Hwey Chen [345] . 494, 510, 540, 574
 Juin-Hwey Chen [217] 240, 246, 247, 249,
 479, 480, 482, 484–487, 494, 534
 Juin-Hwey Chen [291] 348
 Juin-Hwey Chen [232] 247, 414, 423, 439,
 440, 649, 650
 M.J. Cheng [424] 635
 J.C.S. Cheung [277] 318, 320, 371
 J.C.S. Cheung [142] 78
 R.T. Chien [185] 123
 P. S. Chow [380] 582
 Y. C. Chow [80] 51
 C.C. Chu [289] 344
 S. Chua [106] 69, 580
 S. Chua [97] 68, 69, 582
 Soon-Ghee Chua [96] 68, 69
 C.K. Chui [429] 666, 667
 C.K. Chui [430] 663, 666
 L.J. Cimini [116] 75
 Leonard J. Cimini [128] 76, 582
 J. M. Cioffi [380] 582
 A.P. Clark [119] 75
 George C. Clark [171] . 99, 118, 130, 131,
 141, 146
 Ferdinand Classen [375] 582
 Ferdinand Classen [376] 582
 J. Cocke [329] 445, 585, 730
 A. Cohen [433] 663
 P. Combescure [265] 282, 283, 591
 P. Combescure [287] 336
 Daniel J. Constello Jr [174] . 99, 118, 120,
 130, 131, 141
 F.C. Costescu [72] 43
 R. V. Cox [2] 476
 R. V. Cox [1] 476
 R.V. Cox [258] 279
 Richard V. Cox [344] 487
 Richard V. Cox [341] 482
 Richard V. Cox [217] . 240, 246, 247, 249,
 479, 480, 482, 484–487, 494, 534
 Richard V. Cox [291] 348
 Richard V. Cox [300] 357, 362–364
 R.E. Crochiere [352] 553, 601, 733
 R.E. Crochiere [353] 553
 V. Cuperman [463] 776
 Vladimir Cuperman [28] 601

Vladimir Cuperman [224] 246

D

S. Déry [398] 607, 608, 776
Sonia L.Q. Dall'Agnol [350] 536
I. Daubechies [434] 663, 665
G. Davidson [255] 275, 279
Ronald D. Degroat [296] 351
J.R. Deller [19] 602
M. Delprat [286] . 328, 334, 335, 454, 459,
460, 463, 576
E.F. Deprettere [11] . 303, 304, 306, 308,
604
W. Dite [195] 175
Eric M. Dowling [296] 351
J. Durkin [44] 17
W.G. Durtler [69] 43

E

P.M. Ebert [125] 76
R. Edwards [44] 17
P. W. J. van Eetvelt [377] 582
A. El-Jaroudi [421] 634, 666
P. Elias [149] 85
J.A. Erfanian [448] 730
T. Eriksson [270] 293
T. Eriksson [271] 293, 294
E. Erzin [466] 776
D. Esteban [354] 555, 556, 559, 736
B.G. Evans [302] 358, 359
B.G. Evans [279] . 322, 572, 573, 575, 580,
581, 598
B.G. Evans [252] 273
B.G. Evans [303] 360
B.G. Evans [361] 574
B.G. Evans [464] 776
B.G. Evans [452] 740, 812

F

F. Johansen [260] 282
M. Failli [390] 585, 728, 730, 731, 763, 764
G. Falciasacca [145] 81
R.M. Fano [152] 85
N. Farvardin [259] 280
N. Farvardin [237] 261, 278
Stéphane Faudeil [450] 730, 762
K. Fazel [381] 582
K. Fazel [372] 582
K. Feher [67] 43
K. Feher [79] 47, 48
G. Fettweis [372] 582

K. Fischer [265] 282, 283, 591
K.A. Fischer [437] 665, 671, 675
J.L. Flanagan [352] 553, 601, 733
J.L. Flanagan [454] 769
Brian P. Flannery [111] 72, 351–354, 536,
537, 646
G. David Forney [155] 85
G.D. Forney [187] 130, 141, 142
G. David Forney Jr [66] 43
P.M. Fortune [53] 37, 357, 364, 368
L.J. Fransen [240] 262, 264, 265, 269, 618
M. Fratti [293] 349, 354, 355
M. Frullone [145] 81
M. Frullone [144] 81, 380
A. Fudseth [260] 282
K. Fukuda [46] 18
Sadaoki Furui [22] 158

G

C. Galand [354] 555, 556, 559, 736
C.R. Galand [357] 561
Robert G. Gallager [66] 43
J.G. Gardiner [41] 14
William Gardner [318] 404, 406, 412
K. Geher [244] 267
E.B. George [457] 770, 771, 773, 783, 785,
806, 813
E.B. George [458] 770, 771, 785, 806, 813
E.B. George [406] 612
A. Gersho [466] 776
A. Gersho [264] 282, 283
A. Gersho [248] . 273, 287, 288, 494, 789,
792
A. Gersho [223] 246
A. Gersho [255] 275, 279
A. Gersho [347] 494
Allen Gersho [343] 482, 484
Allen Gersho [28] 601
Allen Gersho [230] 246, 397, 482, 485, 497
Allen Gersho [342] 482, 509
Allen Gersho [232] . . 247, 414, 423, 439,
440, 649, 650
I.A. Gerson [316] 398, 402, 426
I.A. Gerson [314] 398, 399, 401, 424
I.A. Gerson [282] 327, 334
I.A. Gerson [315] 398, 399, 401, 424
I.A. Gerson [320] 424, 426
T. Ghiselli-Crippa [421] 634, 666
O. Ghitza [468] 796, 797
R.A. Gibby [124] 76

AUTHOR INDEX

899

Jerry D. Gibson [39] 14, 39
Herbert Gish [215] 223, 273, 276
Alain Glavieux [157]... 85, 444, 582, 584,
726–728, 762
Alain Glavieux [325].....444, 582, 584,
726–728, 762
T.H. Glisson [198] 181
R.M. Golden [454] 769
A. Goldsmith [108].....69, 580
A. Goldsmith [106].....69, 580
A. Goldsmith [97] 68, 69, 582
A. Goldsmith [107].....69, 580
A. Goldsmith [109].....69, 580
Andrea Goldsmith [96]..... 68, 69
S.W. Golomb [189] 131
Gene H. Golub [294].....351
David J. Goodman [141].... 78, 370, 375,
378, 385
G. Gordos [15] 249, 251, 255
D. Gorenstein [181].... 118, 120, 122, 123
D. Gorenstein [162] 85
A.H. Gray [5] 158, 232
Augustine H. Gray [210] 211, 278
R.M. Gray [348] 509, 574
R.M. Gray [248] . 273, 287, 288, 494, 789,
792
P. Grazioso [145].....81
P. Grazioso [144]..... 81, 380
D. Greenwood [43]..... 14, 20, 21, 35
Daniel W. Griffin [225] 246, 607, 610, 733
Yonghai Gu [118] 75, 76
G. Gulak [448]..... 730
A. Le Guyader [265] 282, 283, 591

H

J. Hübner [383].....582
J. Haagen [402] 607, 613, 698
P. Haavisto [333] 454, 458, 459, 463
P. Haavisto [336].....460, 463
R. Hagen [409] 613
J. Hagenauer [449]..... 730, 762
J. Hagenauer [50].....35, 36, 157
Joachim Hagenauer [27] 730
Joachim Hagenauer [326] 444
J.L. Hall [412] 619
R.W. Hamming [164]..... 85, 86
S.L. Hanauer [395] 601–603, 733
T. Hankanen [333].... 454, 458, 459, 463
J.H.L. Hansen [19].....602
H. Harborg [260] 282

C.R.P. Hartmann [159] 85
R. Harun [119].....75
S. Hashimoto [290] 348
H. Hassanein [398] 607, 608, 776
Hisham Hassanein [224].....246
M. Hata [45] 18
Shinji Hayashi [321] 428
S. Haykin [90] 61, 62, 189, 190
J.A. Heller [156] 85
W. Hess [14] 634
D.J. Hiotakakos [410] . 614–616, 685, 690,
692, 698, 705, 706, 732, 812, 843
S. Hirasawa [188] 131, 153
B. Hirosaki [127] 76
A. Hocquenghem [158] 85
Peter Hoehner [445] 729, 730
R. Hoffmann [13] 309
J.N. Holmes [408] 612, 651, 653, 654
W.H. Holmes [218] 240
H. Holtzwarth [194] 175
Masaaki Honda [209]..... 211, 358, 620
Chen Hong [339] 481, 502, 513
T. Honkanen [336] 460, 463
J.J.Y. Huang [256].....276
Johannes B. Huber [374]..... 582
P.M. Huges [247] 269
A.W.F. Huggins [405]..... 612, 795
W.L. Hwang [440] 666

I

T.P. Barnwell III [407] 612
T.P. Barnwell III [400] 607, 612, 613, 651,
653–656, 709, 741, 813
K. Ikeda [319] 412, 605
J. Ikeda [319] 412, 605
M.A. Ireton [283] 328
M.A. Ireton [285] 328, 576
F. Itakura [233] 249
F. Itakura [234] 249
F. Itakura [266] 283, 618
F. Itakura [235] 249
F. Itakura [245] 269
F. Itakura [290] 348
K. Itoh [235] 249
Kenzo Itoh [209] 211, 358, 620
Kenzo Itoh [211] 211

J

J. Lindner [139] 76
I.M. Jacobs [156] 85

Paul Jacobs [318] 404, 406, 412
 A. K. Jain [199] .. 181, 197–199, 275, 276
 William C. Jakes [33] 12, 14, 17
 K. Jarvinen [333] 454, 458, 459, 463
 K. Jarvinen [336] 460, 463
 M.A. Jasiuk [314] 398, 399, 401, 424
 M.A. Jasiuk [282] 327, 334
 M.A. Jasiuk [315] 398, 399, 401, 424
 M.A. Jasiuk [320] 424, 426
 N. Jayant [217] .. 240, 246, 247, 249, 479,
 480, 482, 484–487, 494, 534
 N.S. Jayant [346] 494, 541
 N.S. Jayant [228] 246, 482, 496
 N.S. Jayant [229] 246, 482, 496
 N.S. Jayant [341] 482
 N.S. Jayant [202] 194, 195, 197
 N.S. Jayant [10] .. 166, 172, 175, 179, 181,
 182, 199, 200, 346, 569, 601, 646
 Isabelle Jeanclaude [382] 582
 F. Jelinek [329] 445, 585, 730
 A. Jennings [186] 127, 189, 275, 276
 J.D. Johnston [355] 555, 561
 J.D. Johnston [359] 568
 A. E. Jones [378] 582
 Jr [170] .. 99, 118, 121, 122, 130, 131, 141
 Jr [171] 99, 118, 130, 131, 141, 146
 Jr [187] 130, 141, 142
 Jr. [5] 158, 232
 Biing-Hwang Juang [239] .. 261, 264–266,
 268, 280, 283, 357
 P. Jung [327] 445

K

P. Kabal [241] 265, 267, 268
 P. Kabal [289] 344
 P. Kabal [363] 580
 S. Kadambe [423] 635, 665, 666, 671, 675,
 676
 S. Kaiser [381] 582
 I. Kalet [134] 76, 582
 Y. Kamio [95] 68, 69, 580
 Yuki Yoshi Kamio [74] .. 43, 68, 69, 72, 82,
 580
 Yuki Yoshi Kamio [368] 581
 K.D. Kammeyer [129] 76
 G.S. Kandola [73] 43
 G.S. Kang [240] .. 262, 264, 265, 269, 618
 P. Kapanen [333] 454, 458, 459, 463
 Georges Karam [382] 582
 M. Kasahara [188] 131, 153

A. Kataoka [287] 336
 Akitoshi Kataoka [321] 428
 T. Kawano [46] 18
 W.B. Keijn [32] 601
 P.B. Kenington [71] 43
 R.H. Ketchum [311] 395
 Richard H. Ketchum [349] 527
 R. Kirchherr [265] 282, 283, 591
 A.L. Kirsch [121] 76
 N. Kitawaki [235] 249
 Nobuhiko Kitawaki [209] ... 211, 358, 620
 Nobuhiko Kitawaki [211] 211
 W. Bastiaan Kleijn [300] ... 357, 362–364
 W. Bastiaan Kleijn [349] 527
 W. Bastiaan Kleijn [227] .. 246, 613, 690,
 812
 W.B. Kleijn [311] 395
 W.B. Kleijn [304] 363
 W.B. Kleijn [402] 607, 613, 698
 W.B. Kleijn [409] 613
 A. Klein [453] 762, 763
 J. Knudson [260] 282
 W. Koch [447] 730
 S-N Koh [465] 776
 M.A. Kohler [396] 606
 M.A. Kohler [415] 622
 H.J. Kolb [135] 76
 A.M. Kondoz [302] 358, 359
 A.M. Kondoz [279] ... 322, 572, 573, 575,
 580, 581, 598
 A.M. Kondoz [31] 289, 293, 322, 601
 A.M. Kondoz [252] 273
 A.M. Kondoz [303] 360
 A.M. Kondoz [361] 574
 A.M. Kondoz [464] 776
 A.M. Kondoz [452] 740, 812
 K. Koora [383] 582
 T.H. Koornwinder [428] 663, 666, 667
 D.J. Kraisinsky [311] 395
 Daniel J. Krasinski [349] 527
 E. Woodard Kreamer [419] 622
 E. Woodard Kreamer [417] 622
 P. Kroon [258] 279
 P. Kroon [1] 476
 P. Kroon [287] 336
 P. Kroon [11] 303, 304, 306, 308, 604
 Peter Kroon [300] 357, 362–364
 E. L. Kuan [371] 582, 763
 E.L. Kuan [444] 719
 B. Kull [383] 582

AUTHOR INDEX

901

A. Kumar [466].....776

L

C. Laflamme [306].....370
C. Laflamme [280]322, 438, 530, 575–578,
581

C. Laflamme [261].....282
C. Laflamme [333].....454, 458, 459, 463
C. Laflamme [336].....460, 463
C. Laflamme [281] 322, 576, 579–581, 599
C. Laflamme [322].....428, 518
C. Laflamme [331].....451, 453
C. Laflamme [362].....577, 578, 581
Claude Laflamme [143] 80, 321, 336, 428,
463, 481, 518, 526

Claude Laflamme [332].....451, 453
C. Lamblin [265].....282, 283, 591
C. Lamblin [284].....328
Gordon R. Lang [66].....43
R. Laroia [259].....280
R. Lassalle [132].....76
Vincent K. N. Lau [369].....581
P.A. Laurent [401].....607, 611
H.B. Law [288].....338
Tho Le-Ngoc [118].....75, 76
B. Leach [397].....607, 610
Chong Lee [318].....404, 406, 412
K.Y. Lee [252].....273
W.Y.C. Lee [40].....14, 17
William C. Y. Lee [99].....68, 387
R. Lefebvre [261].....282
A. Lepschy [246].....269
A.H. Levesque [172].....99, 118, 120, 123,
130, 131, 141

S. Levinson [267].....283
C. Li [61].....39, 666
Y. Li [384].....582
Rudolf Lidl [179].....99
P. Lieberman [20].....602
T. H. Liew [366].....581
T. H. Liew [367].....581
T.H. Liew [115].....75, 83
T.H. Liew [114].....75, 581
T.H. Liew [113].....74, 75, 581
Jae S. Lim [225].....246, 607, 610, 733
Shu Lin [174] . 99, 118, 120, 130, 131, 141
Y.-C. Lin [344].....487
Y.C. Lin [217] ... 240, 246, 247, 249, 479,
480, 482, 484–487, 494, 534
Y. Linde [348].....509, 574

J. Linden [270]293
J. Linden [271].....293, 294
S.P. Lloyd [190].....171, 179
S.P. Lloyd [191].....171, 179
Charles F. Van Loan [294].....351
J.L. LoCicero [442]676
J.L. LoCicero [411] ... 614, 685, 692, 719,
812, 843
John H. Lodge [88]61
Fred M. Longstaff [66]43
P. Lupini [463].....776
Peter Lupini [224]246

M

Stefan H. Müller [374]582
P. Mabillean [286]328, 334, 335, 454, 459,
460, 463, 576
P. Mabillean [280]322, 438, 530, 575–578,
581
Malcolm D. Macleod [369]581
X. Maitre [351]549, 562
J. Makhoul [201].....190, 249
J. Makhoul [405].....612, 795
J. Makhoul [236].....257
John Makhoul [182].....123, 190, 191
John Makhoul [215].....223, 273, 276
S. Mallat [435]663, 666, 669
S. Mallat [438]... 666, 668, 669, 671, 812,
840, 841
S. Mallat [440].....666
H. S. Malvar [391].....592, 593, 595
P. Mandarini [379].....582
K. Mano [319]412, 605
J.R.B. De Marca [346]494, 541
J. Roberto B. De Marca [350].....536
J.D. Markel [5]158, 232
John D. Markel [210]211, 278
J.S. Marques [310].....394
J.D. Marvill [71]43
D. Massaloux [265].....282, 283, 591
D. Massaloux [284]328
Dominique Massaloux [143] .80, 321, 336,
428, 463, 481, 518, 526
J.L. Massey [167] . 86, 118, 122, 130, 131,
133, 141
J.L. Massey [153]85
J.L. Massey [166]86
J. Matsumoto [462].....776
Jun Matsumoto [226].....246
Hidehiro Matsuoka [368]581

J. Max [192] 171, 179, 181
 Thomas May [373] 582
 R.J. McAulay [455] 769–772, 806
 R.J. McAulay [459] 770
 R.J. McAulay [403] 609, 776, 779
 R.J. McAulay [461] 776, 795, 796
 R.J. McAulay [399] 607, 609, 776
 R.J. McAulay [460] ... 770, 771, 773, 795,
 796, 806
 A. McCree [406] 612
 A.V. McCree [407] 612
 A.V. McCree [400] 607, 612, 613, 651,
 653–656, 709, 741, 813
 J. P. McGeehan [80] 51
 J.P. McGeehan [91] 64
 J.P. McGeehan [85] 59
 C.A. McGonegal [451] 739
 C.A. McGonegal [426] 651, 653, 654
 M.J. Melchner [217] .. 240, 246, 247, 249,
 479, 480, 482, 484–487, 494, 534
 P. Mermelstein [272] 293
 Heinrich Meyr [375] 582
 Heinrich Meyr [376] 582
 C.A. MGonegal [424] 635
 G.A. Mian [246] 269
 G.A. Miani [293] 349, 354, 355
 A.M. Michelson [172] .. 99, 118, 120, 123,
 130, 131, 141
 A.M. Michelson [160] 85
 S. Miki [319] 412, 605
 Michael L. Moher [88] 61
 J.L. Moncet [289] 344
 N. Morinaga [365] 580
 N. Morinaga [95] 68, 69, 580
 Norihiko Morinaga [74] 43, 68, 69, 72, 82,
 580
 Norihiko Morinaga [368] 581
 Norihiko Morinaga [100] 69
 S. Morissette [286] 328, 334, 335, 454,
 459, 460, 463, 576
 S. Morissette [306] 370
 S. Morissette [280] 322, 438, 530,
 575–578, 581
 S. Morissette [284] 328
 T. Moriya [319] 412, 605
 Takehiro Moriya [321] 428
 C. Mourot [62] 41, 42
 F. Mueller-Roemer [130] 76
 J-M. Muller [320] 424, 426

N

M. NaBhan [327] 445
 Hiromi Nagabucki [211] 211
 M. Naijoh [95] 68, 69, 580
 T. Namekawa [188] 131, 153
 Sanjiv Nanda [141] 78, 370, 375, 378, 385
 K.R. Narayanan [116] 75
 Jon E. Natvig [273] 299, 308
 Harald Niederreiter [179] 99
 Mahesan Niranjana [298] 355
 M. Nishiguchi [462] 776
 Masayuki Nishiguchi [226] 246
 A. R. Nix [80] 51
 A.M. Noll [422] 634
 P. Noll [10] . 166, 172, 175, 179, 181, 182,
 199, 200, 346, 569, 601, 646
 P. Noll [197] 181
 Peter Noll [212] 221
 P.de La Noue [401] 607, 611
 J.M. Nowack [320] 424, 426
 H.J. Nussbaumer [357] 561
 H.J. Nussbaumer [356] 561
 H. Nyquist [76] 46

O

Douglas O'Shaughnessy [17] 158, 602
 H. Ochsner [205] 204
 Elke Offer [326] 444
 E. Ohmori [46] 18
 H. Ohmuro [319] 412, 605
 T. Ojanpare [324] 443, 446
 Y. Okumura [46] 18
 M. Omologo [242] 265
 L.K. Ong [303] 360
 Shinobu Ono [226] 246
 E. Ordentlich [360] 571

P

P. Hoehner [449] 730, 762
 M.D. Paez [198] 181
 K.K. Paliwal [32] 601
 Kuldip K. Paliwal [238] 261, 273, 278,
 279, 286, 357, 629
 P.F. Panter [195] 175
 C. Papanastasiou [467] 776
 Lutz Papke [326] 444
 T. Parks [461] 776, 795, 796
 David Parsons [42] 14, 17, 18
 J.D. Parsons [41] 14
 S. Pasupathy [448] 730

AUTHOR INDEX

903

D. Paul [451] 739
J.W. Paulus [262] 282, 283, 591
D. A. Pearce [98] 68
Peled [126] 76
W.W. Peterson [170] ... 99, 118, 121, 122,
130, 131, 141
W.W. Peterson [161] ... 85, 118, 120, 123
W.W. Peterson [169] 99, 118
N. Phamdo [259] 280
J.W. Picone [442] 676
J.W. Picone [411] 614, 685, 692, 719, 812,
843
S.S. Pietrobon [328] 445
R. Pirhonen [453] 762, 763
G. Plenge [131] 76
V. Pless [175] 99, 122
E.N. Powers [122] 76
V.K. Prabhu [36] 13
Ramjee Prasad [59] 39
William H. Press [111] . 72, 351–354, 536,
537, 646
K. Preuss [137] 76
J.G. Proakis [19] 602
John G. Proakis [48] 21, 25, 50, 586

Q

S.R. Quackenbush [358] 567, 568, 581
T.F. Quatieri [455] 769–772, 806
T.F. Quatieri [459] 770
T.F. Quatieri [461] 776, 795, 796
T.F. Quatieri [399] 607, 609, 776
T.F. Quatieri [460] ... 770, 771, 773, 795,
796, 806
C. Quinquis [265] 282, 283, 591
Shahid U. Qureshi [66] 43

R

R. Rückriem [138] 76
L. Rabiner [267] 283
L.R. Rabiner [420] 634, 666
L.R. Rabiner [451] 739
L.R. Rabiner [424] 635
L.R. Rabiner [6] .. 189–191, 194, 232, 249,
255, 257, 616–618
L.R. Rabiner [426] 651, 653, 654
H. R. Raemer [78] 47
M.D. Anisur Rahham [295] 351
R.P. Ramachandran [241] .. 265, 267, 268
R.P. Ramachandran [250] .. 273, 279, 280
V. Ramamoorthy [228] 246, 482, 496

V. Ramamoorthy [229] 246, 482, 496
K.R. Rao [392] 595
J. Raviv [329] 445, 585, 730
D.K. Ray-Chaudhuri [159] 85
D.K. Ray-Chaudhuri [160] 85
I.S. Reed [163] 86
B. Reiffen [151] 85
Joel R. Remde [9] 244, 301, 603
G. Riccardi [293] 349, 354, 355
O. Rioul [431] 663–665
G. Riva [145] 81
G. Riva [144] 81, 380
P. Robertson [381] 582
Patrick Robertson [446] 730
Patrick Robertson [445] 729, 730
Hermann Rohling [373] 582
A.E. Rosenberg [424] 635
A.E. Rosenberg [427] 651, 653, 654
A.E. Rosenberg [426] 651, 653, 654
Salim Roucos [215] 223, 273, 276
G. Roy [363] 580
L.D. Rudolph [159] 85
M. J. Ruf [381] 582
A. Ruiz [126] 76

S

M. Sabin [461] 776, 795, 796
S. Saito [233] 249
S. Saito [234] 249
R.A. Salami [261] 282
Redwan Ali Salami [276] 315
Redwan Ali Salami [53] . 37, 357, 364, 368
Redwan Ali Salami [280] .. 322, 438, 530,
575–578, 581
Redwan Ali Salami [333] .. 454, 458, 459,
463
Redwan Ali Salami [336] 460, 463
Redwan Ali Salami [254] 273, 275
Redwan Ali Salami [257] 279, 327
Redwan Ali Salami [281] 322, 576,
579–581, 599
Redwan Ali Salami [322] 428, 518
Redwan Ali Salami [143] 80, 321, 336,
428, 463, 481, 518, 526
Redwan Ali Salami [332] 451, 453
Redwan Ali Salami [331] 451, 453
Redwan Ali Salami [200] .. 189, 237, 240,
242–244, 302–304, 306, 308, 322,
325, 329, 358, 362
Redwan Ali Salami [362] ... 577, 578, 581

- Redwan Ali Salami [55] 37, 189–191, 194,
240, 242–244, 257, 301–306, 308,
322, 325, 327, 329, 338, 555, 576
- Redwan Ali Salami [275].....315
- B.R. Saltzberg [123] 76
- M.R. Sambur [426].....651, 653, 654
- S. Sampei [89] 61
- S. Sampei [365].....580
- S. Sampei [95].....68, 69, 580
- Seiichi Sampei [74] 43, 68, 69, 72, 82, 580
- Seiichi Sampei [368] 581
- V.E. Sanchez-Calle [362] ... 577, 578, 581
- Hikmet Sari [382] 582
- Hideichi Sasaoka [74] .. 43, 68, 69, 72, 82,
580
- H.W. Schüssler [136] 76
- R.W. Schafer [6] . 189–191, 194, 232, 249,
255, 257, 616–618
- A. Schmidt-Nielsen [418].....622
- J. Schnitzler [265] 282, 283, 591
- J. Schnitzler [262] 282, 283, 591
- G. Schröder [437] 665, 671, 675
- M.R. Schroeder [412] 619
- Manfred R. Schroeder [216] 237, 238
- Manfred R. Schroeder [16] . 245, 321–323,
501, 604
- P.M. Schultheis [256] 276
- H. Schulze [129].....76
- J. Schur [184].....123
- J. Schur [274].....309
- R. Schwartz [405] 612, 795
- D. Sen [218] 240
- Deep Sen [409] 613
- A. M. Serra [145] 81
- N. Seshadri [250].....273, 279, 280
- R.A. Seymour [288].....338
- C.E. Shannon [49] 35, 157
- S. J. Shepherd [377] 582
- Y. Shoham [360].....571
- Y. Shoham [249].....273
- Y. Shoham [312].....396
- Yair Shoham [409].....613
- S. Singhal [219].....240
- Sharad Singhal [292] 349
- Sharad Singhal [221] 244
- P Sivaprakasapillai [465] 776
- J. Skoeld [453] 762, 763
- J. Skoglung [270] 293
- J. Skoglung [271] 293, 294
- R.J. Sluyter [11]..303, 304, 306, 308, 604
- R.J. Sluyter [12] 309
- B. Smith [196] 175
- M.J.T. Smith [457] ... 770, 771, 773, 783,
785, 806, 813
- M.J.T. Smith [458]770, 771, 785, 806, 813
- K.K.M. So [251] 273
- R. Soheili [361] 574
- N. R. Sollenberger [384].....582
- G. Solomon [163] 86
- M. Sondhi [267].....283
- M.M. Sondhi [250] 273, 279, 280
- Frank K. Soong [239] . 261, 264–266, 268,
280, 283, 357
- S.P. Stapleton [72] 43
- S.P. Stapleton [73].....43
- R.A.J. Steedman [203].....204
- R. Steele [36] 13
- R. Steele [149] 85
- R. Steele [68].....43, 66
- R. Steele [142] 78
- R. Steele [276] 315
- R. Steele [53].....37, 357, 364, 368
- R. Steele [55] 37, 189–191, 194, 240, 242–
244, 257, 301–306, 308, 322, 325,
327, 329, 338, 555, 576
- R. Steele [3] 601
- R. Steele [35] 13
- R. Steele [37] 13
- R. Steele [275].....315
- R. Steele [277] 318, 320, 371
- R. Steele [307].....384, 386–389
- Raymond Steele [94] 68, 580
- Raymond Steele [34] . 12, 29, 33, 727–731
- Raymond Steele [180].....118, 693
- Raymond Steele [93] 68, 69, 82, 580
- J. Stefanov [38]13, 244, 245, 261, 299, 308,
334, 370, 371, 373, 375, 378, 380,
384
- J. Stegmann [265] 282, 283, 591
- J. Stegmann [437] 665, 671, 675
- M. Streeton [62].....41, 42
- H.Y. Su [306] 370
- A.N. Suen [313].....398
- N. Sugamura [245].....269
- N. Sugamura [237] 261, 278
- Y. Sugiyama [188] 131, 153
- R.A. Sukkar [442].....676
- R.A. Sukkar [411]614, 685, 692, 719, 812,
843
- T. Sunaga [89] 61

AUTHOR INDEX

905

R. Suoranta [453] 762, 763
L.M. Supplee [416] 622
L.M. Supplee [396] 606
Consultative Committee for Space Data
Systems [168] 86

T

C. Tai [61] 39, 666
G.Y. Takacs [15] 249, 251, 255
J.D. Tardelli [419] 622
J.D. Tardelli [417] 622
K.A. Teague [397] 607, 610
K.A. Teague [404] 610
Saul A. Teukolsky [111] 72, 351–354, 536,
537, 646
Punya Thitimajshima [157] . 85, 444, 582,
584, 726–728, 762
Timothy Thorpe [213] 222
Uzi Timor [141] ... 78, 370, 375, 378, 385
Jerry V. Tobias [8] 568
Y. Tohkura [290] 348
J.M. Torrance [82] 58, 64, 70, 74
J. M. Torrance [101] 69, 71, 82, 83
J. M. Torrance [102] 69, 72, 82, 83
J. M. Torrance [87] 60, 63–67, 72
J. M. Torrance [103] 69, 82, 83
J. M. Torrance [75] 43, 69, 82, 83
J. Torrance [147] 83
J.M. Torrance [104] . 69, 70, 74, 75, 82, 83
J.M. Torrance [105] 69, 82, 83
J.M. Torrance [146] 83
T. C. Tozer [98] 68
I.M. Trancoso [310] 394
T.E. Tremain [396] 606
T.E. Tremain [415] 622
Thomas E. Tremain [309] .. 392, 396, 637
Thomas E. Tremain [308] .. 392, 602, 604,
606
Thomas Tremain [301] 357, 358, 362, 392
J.M. Tribolet [456] 769
J.M. Tribolet [310] 394
Kwan Truong [406] 612
U. Tuisel [129] 76
F.F. Tzeng [297] 355

U

A. Ubale [264] 282, 283
M. Unser [439] 666
A. Urie [62] 41, 42

V

J. Vainio [333] 454, 458, 459, 463
J. Vainio [336] 460, 463
P. P. Varaiya [109] 69, 580
P. Vary [265] 282, 283, 591
P. Vary [12] 309
P. Vary [13] 309
M. Vetterli [431] 663–665
M. Vetterli [441] 666
William T. Vetterling [111] . 72, 351–354,
536, 537, 646
U. Viaro [246] 269
Emmanuelle Villebrun [445] 729, 730
R. Viswanathan [405] 612, 795
R. Viswanathan [236] 257
V. Viswanathan [406] 612
A. J. Viterbi [51] 35, 157
A.J. Viterbi [154] 85
W.D. Voiers [413] 621
W.D. Voiers [414] 621
F. W. Vook [385] 582

W

Ryuji Wakatsuki [226] 246
B. Walls [404] 610
J.F. Wand [313] 398
D. Wang [263] 282
Hong Shen Wang [340] 481
S. Wang [223] 246
I. Wassell [55] 37, 189–191, 194, 240, 242–
244, 257, 301–306, 308, 322, 325,
327, 329, 338, 555, 576
S.A. Webber [352] 553, 601, 733
S.B. Weinstein [125] 76
Vanoy C. Welch [309] 392, 396, 637
Vanoy Welch [301] 357, 358, 362, 392
E.J. Weldon [170] . 99, 118, 121, 122, 130,
131, 141
R.J. Wilkinson [71] 43
R.J. Wilkinson [70] 43
T. A. Wilkinson [378] 582
J. Williams [277] 318, 320, 371
J. Williams [307] 384, 386–389
E.H. Winter [320] 424, 426
C. H. Wong [371] 582, 763
C. H. Wong [366] 581
C.H. Wong [115] 75, 83
C.H. Wong [112] 74, 83, 581
C.H. Wong [114] 75, 581
C.H. Wong [113] 74, 75, 581

K. H. H. Wong [54] 37, 189, 191
K. H. J. Wong [55] 37, 189–191, 194, 240,
242–244, 257, 301–306, 308, 322,
325, 327, 329, 338, 555, 576
K.H.H. Wong [149] 85
K.H.H. Wong [177] 99, 118, 130, 141
J. Woodard [305] 369
J.P. Woodard [323] 443, 447, 449, 450, 582
J.P. Woodard [443] 719
J.P. Woodard [386] 582
J.P. Woodard [299] 356, 485
J.P. Woodard [110] 71, 508, 533
Jason P. Woodard [52] . 37, 331, 369, 465,
533, 542
J.M. Wozencraft [150] 85
J.M. Wozencraft [151] 85
A.S. Wright [69] 43
J. Wu [117] 75, 76
C. W. Wyatt–Millington [377] 582

X

C.S. Xydeas [410] 614–616, 685, 690, 692,
698, 705, 706, 732, 812, 843
C.S. Xydeas [283] 328
C.S. Xydeas [467] 776
C.S. Xydeas [285] 328, 576
C.S. Xydeas [251] 273

Y

H. Yang [465] 776
T.C. Yao [313] 398
B.L. Yeap [443] 719
M. S. Yee [367] 581
S. Yeldner [464] 776
S. Yeldner [452] 740, 812
P. Yip [392] 595
M. Yong [255] 275, 279
Kai-Bor Yu [295] 351

Z

H. Zarrinkoub [272] 293
K.A. Zeger [347] 494
S. Zeisberg [383] 582
R. Zelinski [197] 181
Rabiner Zelinski [212] 221
Jian Zhang [340] 481
C. Zheng [61] 39, 666
S. Zhong [438] ... 666, 668, 669, 671, 812,
840, 841
N. Zierler [181] 118, 120, 122, 123
N. Zierler [162] 85