

# Principles of Speech Coding

Tokunbo Ogunfunmi  
Madihally Narasimha

TECHNISCHE  
INFORMATIONSBIBLIOTHEK  
UNIVERSITÄTSBIBLIOTHEK  
HANNOVER



**CRC Press**  
Taylor & Francis Group  
Boca Raton London New York

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business

---

# Contents

---

Foreword	xv
Preface	xvii
Acknowledgments	xxi
Authors	xxiii
<b>1. Introduction to Speech Coding</b>	<b>1</b>
1.1 Introduction	1
1.2 Speech Signals	1
1.3 Characteristics of Speech Signals	3
1.4 Modeling of Speech	7
1.5 Speech Analysis	8
1.6 Speech Coding	9
1.6.1 A Very Brief History of Speech Coding	9
1.6.2 Major Classification of Speech Coders	10
1.6.3 Speech Coding Standards	12
1.7 Varieties of Speech Coders	13
1.7.1 Varieties of Waveform Speech Coders	13
1.7.2 Varieties of Parametric (Analysis-by-Synthesis) Speech Coders	14
1.8 Measuring Speech Quality	14
1.8.1 Mean Opinion Score	14
1.8.2 Perceptual Evaluation of Speech Quality	15
1.8.3 Enhanced Modified Bark Spectral Distance	15
1.8.4 Diagnostic Rhyme Test	15
1.8.5 Diagnostic Acceptability Measure	15
1.8.6 E-Model	16
1.9 Communication Networks and Speech Coding	16
1.10 Performance Issues in Speech Communication Systems	17
1.10.1 Speech Quality	17
1.10.2 Communication Delay	18
1.10.3 Computational Complexity	18
1.10.4 Power Consumption	18
1.10.5 Robustness to Noise	18
1.10.6 Robustness to Packet Losses (for Packet-Switched Networks)	18
1.11 Summary of Speech Coding Standards	18
1.12 Summary	23
Exercise Problems	24

References . . . . .	24
Bibliography . . . . .	25
<b>2. Fundamentals of DSP for Speech Processing . . . . .</b>	<b>27</b>
2.1 Introduction to LTI Systems . . . . .	27
2.1.1 Linearity . . . . .	27
2.1.2 Time Invariance . . . . .	27
2.1.3 Representation Using Impulse Response . . . . .	28
2.1.4 Representation of Any Continuous-Time (CT) Signal . . . . .	29
2.1.5 Convolution . . . . .	29
2.1.6 Differential Equation Models . . . . .	30
2.2 Review of Digital Signal Processing . . . . .	31
2.2.1 Sampling . . . . .	31
2.2.2 Shifted Unit Pulse: $\delta(n - k)$ . . . . .	31
2.2.3 Representation of Any DT Signal . . . . .	32
2.2.4 Introduction to Z Transforms . . . . .	36
2.2.5 Fourier Transform, Discrete Fourier Transform . . . . .	37
2.2.6 Digital Filter Structures . . . . .	38
2.3 Review of Stochastic Signal Processing . . . . .	41
2.3.1 Power Spectral Density . . . . .	43
2.4 Response of a Linear System to a Stochastic Process Input . . . . .	44
2.5 Windowing . . . . .	45
2.6 AR Models for Speech Signals, Yule–Walker Equations . . . . .	47
2.7 Short-Term Frequency (or Fourier) Transform and Cepstrum . . . . .	47
2.7.1 Short-Term Frequency Transform (STFT) . . . . .	47
2.7.2 The Cepstrum . . . . .	48
2.8 Periodograms . . . . .	48
2.9 Spectral Envelope Determination for Speech Signals . . . . .	49
2.10 Voiced/Unvoiced Classification of Speech Signals . . . . .	51
2.10.1 Time-Domain Methods . . . . .	51
2.10.1.1 Periodic Similarity . . . . .	51
2.10.1.2 Frame Energy . . . . .	52
2.10.1.3 Pre-Emphasized Energy Ratio . . . . .	52
2.10.1.4 Low- to Full-Band Energy Ratio . . . . .	52
2.10.1.5 Zero Crossing . . . . .	52
2.10.1.6 Prediction Gain . . . . .	53
2.10.1.7 Peakiness of Speech . . . . .	53
2.10.1.8 Spectrum Tilt . . . . .	53
2.10.2 Frequency-Domain Methods . . . . .	53
2.10.3 Voiced/Unvoiced Decision Making . . . . .	54
2.11 Pitch Period Estimation Methods . . . . .	54
2.12 Summary . . . . .	57
Exercise Problems . . . . .	57
References . . . . .	58
Bibliography . . . . .	59

<b>3. Sampling Theory</b> . . . . .	61
3.1 Introduction . . . . .	61
3.2 Nyquist Sampling Theorem . . . . .	61
3.3 Reconstruction of the Original Signal: Interpolation Filters . . . . .	62
3.4 Practical Reconstruction . . . . .	63
3.5 Aliasing and In-Band Distortion . . . . .	64
3.6 Effect of Sampling Clock Jitter . . . . .	65
3.7 Sampling and Reconstruction of Random Signals . . . . .	67
3.8 Summary . . . . .	68
Exercise Problems . . . . .	68
Reference . . . . .	68
Bibliography . . . . .	68
<b>4. Waveform Coding and Quantization</b> . . . . .	69
4.1 Introduction . . . . .	69
4.2 Quantization . . . . .	69
4.3 Quantizer Performance Evaluation . . . . .	70
4.4 Quantizer Transfer Function . . . . .	71
4.5 Quantizer Performance under No-Overload Conditions . . . . .	72
4.6 Uniform Quantizer . . . . .	73
4.7 Nonuniform Quantizer . . . . .	75
4.7.1 Nonuniform Quantizer Implementation Methods . . . . .	76
4.7.2 Nonuniform Quantizer Performance . . . . .	77
4.8 Logarithmic Companding . . . . .	80
4.8.1 Approximations to Logarithmic Companding . . . . .	80
4.8.1.1 $\mu$ -Law (Continuous Version) . . . . .	81
4.8.1.2 $A$ -Law (Continuous Version) . . . . .	82
4.8.2 Companding Advantage . . . . .	84
4.9 Segmented Companding Laws . . . . .	85
4.9.1 Segmented Approximation to the Continuous $\mu$ -Law and $A$ -Law Curves . . . . .	86
4.10 ITU G.711 $\mu$ -Law and $A$ -Law PCM Standards . . . . .	91
4.10.1 Conversion between Linear and Companded Codes . . . . .	91
4.10.1.1 Linear to $\mu$ -Law Conversion . . . . .	92
4.10.1.2 $\mu$ -Law to Linear Code Conversion . . . . .	93
4.10.1.3 Linear to $A$ -Law Conversion . . . . .	94
4.10.1.4 $A$ -Law to Linear Conversion . . . . .	95
4.11 Optimum Quantization . . . . .	95
4.11.1 Closed Form Solution for the Optimum Companding Characteristics . . . . .	96
4.11.2 Lloyd–Max Quantizer . . . . .	97
4.12 Adaptive Quantization . . . . .	99
4.13 Summary . . . . .	101
Exercise Problems . . . . .	102

References . . . . .	108
Bibliography . . . . .	108
<b>5. Differential Coding . . . . .</b>	<b>111</b>
5.1 Introduction . . . . .	111
5.2 Closed-Loop Differential Quantizer . . . . .	112
5.3 Generalization to Predictive Coding . . . . .	113
5.3.1 Optimum Closed-Loop Predictor . . . . .	115
5.3.2 Adaptive Prediction . . . . .	117
5.4 ITU G.726 ADPCM Algorithm . . . . .	118
5.4.1 Adaptive Quantizer . . . . .	119
5.4.1.1 Quantizer Scale Factor Adaption . . . . .	121
5.4.1.2 Quantizer Adaption Speed Control . . . . .	121
5.4.2 Predictor Structures and Adaption . . . . .	123
5.5 Linear Deltamodulation . . . . .	125
5.5.1 Optimum 1-Bit Quantizer . . . . .	126
5.5.2 Optimum Step Size and SNR . . . . .	126
5.5.2.1 Special Cases . . . . .	128
5.5.2.2 SNR for Sinusoidal Inputs with Perfect Integration . . . . .	128
5.6 Adaptive Deltamodulation . . . . .	129
5.7 Summary . . . . .	130
Exercise Problems . . . . .	131
Reference . . . . .	133
Bibliography . . . . .	133
<b>6. Linear Prediction . . . . .</b>	<b>135</b>
6.1 Introduction . . . . .	135
6.1.1 Linear Prediction Theory and Wiener Filters . . . . .	135
6.2 Properties of the Autocorrelation Matrix, $R$ . . . . .	138
6.3 Forward Linear Prediction . . . . .	139
6.4 Relation between Linear Prediction and AR Modeling . . . . .	142
6.5 Augmented Wiener–Hopf Equations for Forward Prediction . . . . .	142
6.6 Backward Linear Prediction . . . . .	143
6.7 Backward Prediction-Error Filter . . . . .	146
6.8 Augmented Wiener–Hopf Equations for Backward Prediction . . . . .	147
6.9 Relation between Backward and Forward Predictors . . . . .	147
6.10 Levinson–Durbin Recursion . . . . .	150
6.10.1 L-D Algorithm . . . . .	150
6.10.2 Forward Linear Prediction . . . . .	151
6.10.3 Backward Linear Prediction . . . . .	151
6.10.4 Inverse L-D Algorithm . . . . .	160
6.10.5 Summary of L-D Recursion . . . . .	162
6.11 Summary . . . . .	162

Exercise Problems	162
References	164
Bibliography	164
<b>7. Linear Predictive Coding</b>	<b>165</b>
7.1 Introduction	165
7.2 Linear Predictive Coding	165
7.2.1 Excitation Source Models	166
7.3 LPC-10 Federal Standard	171
7.3.1 Encoder	171
7.3.2 LPC Decoder	172
7.3.3 FS-1015 Speech Coder	172
7.4 Introduction to CELP-Based Coders	174
7.4.1 Perceptual Error Weighting	179
7.4.2 Pitch Estimation	180
7.4.3 Closed-Loop Pitch Search (Adaptive Codebook Search)	180
7.5 Summary	182
Exercise Problems	182
References	183
Bibliography	183
<b>8. Vector Quantization for Speech Coding Applications</b>	<b>185</b>
8.1 Introduction	185
8.2 Review of Scalar Quantization	186
8.3 Vector Quantization	186
8.3.1 The Overall Distortion Measure	188
8.3.2 Distortion Measures	190
8.3.3 Codebook Design	192
8.4 Lloyd's Algorithm for Vector Quantizer Design	192
8.4.1 Splitting Method	195
8.5 The Linde–Buzo–Gray Algorithm	195
8.6 Popular Search Algorithms for VQ Quantizer Design	196
8.6.1 Full Search VQ	196
8.6.2 Binary Search VQ	197
8.7 Other Suboptimal Algorithms for VQ Quantizer Design	198
8.7.1 Multistage VQ	198
8.7.2 Split VQ	201
8.7.3 Conjugate VQ	201
8.7.4 Predictive VQ	202
8.7.5 Adaptive VQ	202
8.8 Applications in Standards	203
8.9 Summary	203
Exercise Problems	204
References	204
Bibliography	205

<b>9. Analysis-by-Synthesis Coding of Speech</b> . . . . .	207
9.1 Introduction . . . . .	207
9.2 CELP AbS Structure . . . . .	208
9.3 Case Study Example: FS 1016 CELP Coder . . . . .	210
9.4 Case Study Example: ITU-T G.729/729A Speech Coder . . . . .	216
9.4.1 The ITU G.729/G.729A Speech Encoder . . . . .	217
9.4.1.1 The ITU G.729 Encoder Details . . . . .	220
9.4.1.2 Quantization of the Gains . . . . .	235
9.4.2 The ITU G.729/G.729A Speech Decoder . . . . .	235
9.4.2.1 The ITU G.729 Decoder Details . . . . .	235
9.4.2.2 Long-Term Postfilter . . . . .	237
9.4.2.3 Short-Term Postfilter . . . . .	238
9.4.2.4 High-Pass Filtering and Upscaling . . . . .	238
9.4.2.5 Tilt Compensation . . . . .	238
9.4.2.6 Adaptive Gain Control . . . . .	239
9.5 Summary . . . . .	239
Exercise Problems . . . . .	240
References . . . . .	241
Bibliography . . . . .	241
<b>10. Internet Low-Bit-Rate Coder</b> . . . . .	243
10.1 Introduction . . . . .	243
10.2 Internet Low-Bit-Rate Codec . . . . .	244
10.2.1 Structure . . . . .	244
10.2.2 Advantages . . . . .	245
10.2.3 Algorithm . . . . .	245
10.2.4 CELP Coders versus iLBC . . . . .	246
10.3 iLBC's Encoding Process . . . . .	247
10.4 iLBC's Decoding Process . . . . .	251
10.5 iLBC's PLC Techniques . . . . .	254
10.6 iLBC's Enhancement Techniques . . . . .	255
10.6.1 Outline of Enhancer . . . . .	256
10.7 iLBC's Synthesis and Postfiltering . . . . .	259
10.8 MATLAB® Signal Processing Blockset iLBC Demo Model . . . . .	259
10.9 PESQ . . . . .	267
10.10 Evolution from PSQM/PSQM+TO PESQ . . . . .	267
10.10.1 PSQM+ . . . . .	268
10.11 PESQ Algorithm . . . . .	268
10.12 PESQ Applications . . . . .	271
10.13 Summary . . . . .	271
Exercise Problems . . . . .	272
References . . . . .	272
Bibliography . . . . .	273

<b>11. Signal Processing in VoIP Systems</b> . . . . .	275
11.1 Introduction . . . . .	275
11.2 PSTN and VoIP Networks . . . . .	275
11.3 Effect of Delay on the Perceived Speech Quality . . . . .	276
11.4 Line Echo Canceler . . . . .	278
11.4.1 Adaptive Filter . . . . .	279
11.4.2 Double-Talk Detector . . . . .	281
11.4.3 Nonlinear Processor . . . . .	282
11.4.4 Comfort Noise Generator . . . . .	282
11.5 Acoustic Echo Canceler . . . . .	283
11.6 Jitter Buffers . . . . .	285
11.7 Clock Skew . . . . .	288
11.8 Packet Loss Recovery Methods . . . . .	289
11.8.1 Transmitter-Based FEC Techniques . . . . .	290
11.8.2 Receiver-Based PLC Algorithms . . . . .	291
11.9 Summary . . . . .	293
Bibliography . . . . .	294
<b>12. Real-Time DSP Implementation of ITU-T G.729/A</b>	
<b>Speech Coder</b> . . . . .	295
12.1 Introduction . . . . .	295
12.2 ITU-T G.729/A Speech Coding Standard . . . . .	296
12.3 TI TMS320C6X DSP Processors . . . . .	297
12.4 TI's RF and DSP Algorithm Standard . . . . .	300
12.5 G.729/A on RF3 on the TI C6X DSP . . . . .	301
12.5.1 IALG Interface . . . . .	302
12.5.2 ALGRF . . . . .	303
12.6 Running the RF3 Example on EVM . . . . .	304
12.7 RF3 Resource Requirements . . . . .	304
12.7.1 RF3 Memory Requirements . . . . .	304
12.7.2 RF3 Clock Cycle Requirements . . . . .	306
12.8 Details of Our Implementation . . . . .	306
12.8.1 Adapting, Building, and Running the G.729/A Code . . . . .	306
12.8.2 Defining the Data Type Sizes for the Vocoder . . . . .	307
12.8.3 Early Development Using Microsoft Visual Studio . . . . .	308
12.8.4 Microsoft Visual Studio Encoder Project . . . . .	311
12.8.5 Microsoft Visual Studio Decoder Project . . . . .	312
12.8.5.1 Comparing Test Vectors . . . . .	312
12.8.5.2 Measuring Performance Timing on Microsoft Visual Studio . . . . .	313
12.8.5.3 Automating the Test Vector Comparisons on Windows . . . . .	313
12.9 Migrating ITU-T G.729/A to RF3 and the EVM . . . . .	314



12.9.1	Creating a New Application . . . . .	314
12.9.1.1	Adapting the Vocoder Library Files for the EVM . . . . .	315
12.9.1.2	G.729/A Application for RF3 . . . . .	315
12.9.1.3	algG729A and algInvG729A (Function Wrappers) . . . . .	315
12.9.1.4	appModules . . . . .	315
12.9.1.5	C67xEMV_RF3 (Application Project) . . . . .	317
12.9.1.6	Building the G.729/A Vocoder Application . . . . .	318
12.10	Optimizing G.729/A for Real-Time Execution on the EVM . . . . .	318
12.10.1	Project Settings . . . . .	318
12.10.2	DSP/BIOS Settings for Optimization . . . . .	319
12.10.3	Code Changes for Optimization . . . . .	320
12.11	Real-Time Performance for Two Channels . . . . .	322
12.11.1	Memory Requirements for G.729/A . . . . .	322
12.11.2	Clock Cycle Requirements for G.729/A . . . . .	323
12.11.3	Resource Requirements Summary . . . . .	324
12.12	Checking the Test Vectors on the EVM . . . . .	324
12.13	Going beyond a Two-Channel Implementation . . . . .	325
12.13.1	Adding Channels . . . . .	326
12.13.2	DSP/BIOS Changes for Adding Channels . . . . .	326
12.13.2.1	Source Code Changes for Adding Channels . . . . .	327
12.13.3	Running Seven Channels of the Vocoder on the EVM . . . . .	327
12.13.4	Getting Eight Channels on the G.729/A Application . . . . .	328
12.13.5	Going beyond Eight Channels in the G.729/A Application . . . . .	329
12.13.5.1	Profiling the Vocoder from the Top Level . . . . .	329
12.13.5.2	Profiling the Encoder . . . . .	330
12.13.5.3	Profiling ACELP_Code_A . . . . .	330
12.14	Conclusions . . . . .	332
	References . . . . .	332
	Bibliography . . . . .	333
<b>13.</b>	<b>Conclusions and Future Directions for Speech Coding . . . . .</b>	<b>335</b>
13.1	Summary . . . . .	335
13.2	Future Directions for Speech Research . . . . .	336
	References . . . . .	339
	Bibliography . . . . .	340
<b>Index . . . . .</b>		<b>341</b>