Contents

1 Introduction 1
1.1 Lower Limit of Bit-Rate 3
1.1.1 Using Text Accompanying Speech 4
1.2 Vocoder Framework 6
1.3 Clustered Codebook to Continuous Codebook 9
1.3.1 Clustered Segment Codebook 9
1.3.2 Vector Quantization Performance Trends and Limits 10
1.3.3 Random Segment Codebooks 10
1.3.4 Vector to Segment Quantization Performance Retention 11
1.3.5 A Converging Viewpoint 12
1.3.5.1 Reasoning I 12
1.3.5.2 Reasoning II 15
1.4 Speech-to-Speech Synthesis by Unit-Selection 18
1.5 Alternate Perspectives for Ultra Low Bit-Rate Speech Coding . . . 21
1.6 Applications of Ultra Low Bit-Rate Speech Coding 23
1.7 Organization of the Book 25

2 Ultra Low Bit-Rate Coders 27
2.1 Vector and Matrix Quantization 29
2.2 Segment Vocoders 32
2.2.1 Automatic Segmentation 34
2.2.1.1 Spectral Transition Measure 35
2.2.1.2 Maximum-Likelihood Segmentation 36
2.2.1.3 ML Segmentation: Duration Constrained (ML(DC)) 38
2.2.1.4 ML Segmentation: A Generalized Basis 40
2.2.1.5 Syllable-like units and other segmentations 40
2.2.1.6 Temporal Decomposition 41
2.2.2 Segment Quantization 42
2.2.3 Joint Segmentation Quantization 44
2.2.3.1 Basic framework 44

2.2.3.2 Shiraki and Honda Variable-Length Segment Quantization . . . 45
2.2.3.3 2-Level DP Framework for Joint Segmentation and Quantization . . . 46
2.2.3.4 One-Pass DP Algorithm . . . 48
2.2.3.5 Phoneme Recognition and Phonetic Vocoders . . . 48
2.2.4 Segment Codebook . . . 50
2.2.4.1 Template Segment Codebooks . . . 51
2.2.4.2 HMM Segment Codebook . . . 53
2.2.5 Duration Modification . . . 53
2.2.6 Residual Parameterization and Quantization . . . 55
2.2.7 Synthesis . . . 56
2.3 R/D Optimal Linear Prediction . . . 57
2.3.1 Prandoni and Vetterli R/D Optimal Linear Prediction . . . 57
2.3.2 Variable-to-Variable Length Vector Quantization . . . 60
2.3.3 Multigrams Quantization . . . 61
2.3.4 Distortion Constrained Segmentation . . . 61
2.4 HMM Based Recognition-Synthesis Paradigm . . . 62
2.4.1 HTS Based Framework . . . 63
2.4.2 Speaker Adaptive HMM Recognition-Synthesis . . . 64
2.4.3 Ergodic HMM Framework . . . 65
2.4.4 Ismail and Ponting HMM Based Vocoders . . . 66
2.4.5 Formant Trajectory Model Based Recognition-Synthesis . . . 67
2.5 ALISP Units and Refinements . . . 67
2.5.1 Basic ALISP Framework . . . 68
2.5.2 Re-segmented Long Synthesis Units . . . 69
2.5.3 Short Synthesis Units by Dynamic Unit Selection . . . 70
2.5.4 Pre-selection of Units . . . 70
2.5.5 Noise Robustness . . . 71
2.6 Speaker Adaptation in Phonetic Vocoders . . . 71
2.7 Unit-Selection Paradigms . . . 72
2.8 Performance Measures for Segment Quantization . . . 75

3 Unit Selection Framework . . . 79
3.1 Lee-Cox Single-Frame Unit Selection Quantization . . . 81
3.1.1 An Alternate ‘5 ms Segment’ Single-Frame Unit-Selection Algorithm . . . 82
3.2 Lee-Cox Segmental Unit Selection Quantization . . . 83
3.3 Run-Length Coding and Effective Bit-Rate . . . 88
3.4 Sub-optimality of Lee-Cox Segmental Unit-Selection Algorithm . . . 90

4 Unified and Optimal Unit-Selection Framework 91
4.1 Unified Unit-Selection Framework 92
4.1.1 Proposed One-Pass DP Algorithm 94
4.1.1.1 Comparison with Lee and Cox Single-Frame and Segmental Unit-Selection 96
4.1.2 Experiments and Results 98
4.2 Comparison with Lee and Cox Suboptimal Segmental Unit Selection 100
4.3 Comparison with VQ, MQ, VLSQ 103
4.3.1 Experiments and Results 104
4.4 Conclusions 107

5 Optimality and Complexity Considerations 109
5.1 Complexity of 1-Pass DP Optimal Unit-Selection Algorithm 110
5.1.1 Local Distance Calculations Cost 111
5.1.2 Within-Unit Recursion Cost 111
5.1.3 Cross-Unit Recursion Cost 111
5.2 Complexity of Lee-Cox Segmental Unit-Selection Algorithm 112
5.2.1 Distance Calculations 113
5.2.2 Recursion Calculations 114
5.3 Comparison of 1-Pass DP and Lee-Cox Segmental Unit-Selection 114
5.4 Optimality-Complexity Tradeoff 115
5.5 Proposed n-Best Lattice Search 116
5.6 Conclusions 123

6 No Residual Transmission: Joint Spectral-Residual Quantization 125
6.1 Joint Spectral-Residual Quantization in Lee-Cox Unit-Selection Framework 127
6.1.1 Joint Spectral-Residual Quantization 129
6.1.2 Experiments and Results 133
6.2 Joint Spectral-Residual Quantization in an Optimal Unit-Selection Framework 136
6.2.1 Joint Spectral-Residual Quantization in 1-Pass DP Framework 137
6.2.2 Experiments and Results 140
6.3 Conclusions 143

References 145

http://www.springer.com/978-1-4939-1340-4

Ultra Low Bit-Rate Speech Coding
Ramasubramanian, V.; Doddala, H.
2015, VII, 152 p. 60 illus., 56 illus. in color., Softcover
ISBN: 978-1-4939-1340-4