

Contents

Acknowledgements	iv
Abstract	v
Contents	vi
List of Figures	x
List of Tables.	xiii
List of Acronyms	xiv
1 Introduction	
1.1 Background	1
1.2 Perceptual Audio Coding	1
1.3 Coding in the Multimedia Scenario	2
1.4 Thesis Contributions	3
1.5 Organization of the thesis	3
2 Psychoacoustics and the Human Hearing	4
2.1 The Human Ear	6
2.1.1 Outer Middle and Inner Ear	6
2.1.2 The Basilar Membrane	6
2.1.3 Critical Bands	7
2.1.4 Absolute Threshold of Hearing	9
2.2 Masking	10
2.2.1 Simultaneous Masking	13
2.2.1.1 Noise Masking Tone	13
2.2.1.2 Tone Masking Noise	14
2.2.1.3 Noise Masking Noise	15

2.2.2	Spread of Masking	17
2.2.3	Temporal Masking	18
2.3	Perceptual Entropy	18
2.3.1	Critical Band Analysis.	19
2.4	Psychoacoustic Models	20
2.4.1	MPEG Psychoacoustic Model – I	21
2.4.2	MPEG Psychoacoustic Model – II	26
3	Transform and Subband Audio Coding	29
3.1	Transform Coding	29
3.1.1	Block Transforms	29
3.1.2	Basic Transform Coding Structure	30
3.1.3	Transform Coding with The DFT	32
3.2	Filter Banks and Subband Coding	34
3.2.1	A Basic Subband Coding Frame Work	35
3.2.2	Conditions for Perfect Reconstruction	37
3.2.3	Filter Bank Designs for Audio Coding	37
3.2.4	Transform Vs Subband Coding	39
3.2.5	The Modified Discrete Cosine Transform (MDCT)	40
3.3	Quantization and Coding	42
3.3.1	Uniform Quantization	42
3.3.2	Non Uniform Quantization	43
3.3.3	Entropy Coding	43
3.3.3.1	Huffman Coding	44
3.4	Bit Allocation in Audio Source coding.	44
3.5	Audio Quality Measurement	44
3.5.1	Subjective Measures.	45
3.5.2	Objective Measures.	47
4	Efficient Audio Coding with Optimized Subbands	48
4.1	Quantization Optimization	48
4.1.1	General	48
4.1.2	Motivation for Optimized Subbands	49

4.2	Non Uniform Non Critical (NUNC) Subbands	51
4.2.1	Design Considerations	52
4.2.1.1	Critical Band Boundaries	52
4.2.1.2	Number of Samples in a NUNC Band	53
4.2.1.3	Number of NUNC Bands in a Frame	53
4.2.2	Prominent Spectral Peaks (PSPs)	53
4.2.3	NUNC Bands Formulation Algorithm	54
4.2.4	The Class Analysis Algorithm	60
4.2.5	NUNC Configurations	61
4.3	Application to MPEG-2/4 AAC Encoding	64
4.3.1	Quantization in MPEG-2/4 AAC.	65
4.3.2	Recent Improvements	67
4.3.3	Incorporating NUNC bands in an MPEG-4 Encoder.	68
4.4	Simulation and Performance Analysis.	68
4.4.1	Test Material	68
4.4.2	Subjective Listening Tests.	70
4.4.3	Objective Analysis	71
5	Perceptual Coding with Coarse Signal Representation	88
5.1	Coarse Signal Representation (CSR).	88
5.2	Perceptual Motivation for CSR	89
5.3	CSR Coding Criteria.	90
5.3.1	CSR Coding for High Frequencies.	91
5.3.2	CSR Coding for Low Frequencies.	92
5.4	Codec Structure for CSR Coding	94
5.4.1	Bit Allocation and Quantization	95
5.4.1.1	Bit Demand Computation	98
5.4.1.2	Bit Allocation.	96
5.4.1.3	Quantization	97

5.5	Simulation and Performance Analysis	97
5.5.1	Subjective Listening Tests.	98
5.5.2	Objective Analysis	102
6	Integrated Speech and Audio Coding	111
6.1	Introduction	111
6.2	Related Work	112
6.2.1	Merged Transform and Linear Predictive Coding.	112
6.2.2	Hybrid Switched Coding	113
6.2.3	Layered Coding	113
6.2.4	Need for Integrated Coding	114
6.3	Signal Classification for Codec Selector Algorithm.	115
6.3.1	Speech/Non Speech Discrimination	115
6.3.2	Audio Signal Sub Classification	116
6.3.2.1	Transient/Stationary Signal Classification	116
6.3.2.2	Sub Classification of Stationary Signals	121
6.4	The Codec Selector Algorithm	123
6.4.1	Encoder Structure	123
6.4.2	Codec Selection Decision	124
6.4.2.1	Choice Between DFT and MDCT	125
6.5	Performance Analysis	131
6.5.1	Test Material	131
6.5.2	Subjective Listening Tests	131
7	Conclusions	133
7.1	Thesis Summary	133
7.2	Future Research Directions	134
7.2.1	NUNC Coding Paradigm	134
7.2.2	CSR Coding	135
7.2.3	Integrated Speech and Audio Coding	135
	Publications	136
	Bibliography	138