

Contents

Abstract	iii
Resumen	v
Resum	vii
Acknowledgements	ix
List of symbols	xxi
Abbreviations and Acronyms	xxiii
1 Introduction and Scope	1
1.1 Background	3
1.2 Motivation	5
1.3 Scope of the Thesis	6
1.4 Organization of the Thesis	6
I Source Separation	9
2 Sound Source Separation	11
2.1 Introduction	13
2.2 Sources and Mixtures	16
2.2.1 Audio Sources	16
2.2.2 Audio Mixtures	16
2.3 Mixing models	19
2.3.1 Instantaneous Model	20
2.3.2 Anechoic Model	21
2.3.3 Convolutional Model	22
2.3.4 Noisy models	22
2.4 Source Separation Tasks and Approaches	23
2.4.1 Problem Classification	23
2.4.2 Joint and Staged Approaches	24
2.5 Underdetermined Source Separation	27
2.5.1 Additive Expansions	28
2.5.2 Basis decompositions	28
2.5.3 The Short-Time Fourier Transform (STFT)	30
2.5.4 Sparsity	33
2.5.5 Estimation of the Sources	35
2.5.6 Time-Frequency Masking	38
2.6 Algorithms for Stereo Mixtures	40
2.6.1 DUET	40
2.6.2 Panning Index Window (PIW)	41
2.6.3 ADReSS	42
2.7 Monaural Separation	43

2.7.1	Non-Negative Matrix Factorization	43
2.7.2	Computational Auditory Scene Analysis	45
2.8	Performance measurement	45
2.9	Conclusion	48
3	Multi-Level Thresholding Separation	49
3.1	Introduction	51
3.2	Separation of instantaneous mixtures	53
3.2.1	Pan Map	54
3.2.2	Weighted histograms	56
3.2.3	Thresholding	58
3.2.4	Efficient implementation	64
3.2.5	Estimation of the sources	65
3.2.6	Computational Complexity	65
3.2.7	Performance Evaluation	66
3.3	Separation in Real Environments	71
3.3.1	DOA map	72
3.3.2	Coherence-Based Selection	73
3.3.3	Histogram formation	76
3.3.4	Multi-Level Thresholding and Source Estimation	77
3.3.5	Performance Evaluation	78
3.4	Conclusion	82
4	Post-Processing Separated Sources	85
4.1	Introduction	87
4.2	Residual Extraction by Energy Masking	88
4.2.1	Energy-Normalized Source to Interference Ratio	88
4.2.2	Binary Masks for Residuals Removal	90
4.2.3	Experiments	90
4.3	Neighborhood-Based Reassignment	94
4.3.1	Cluster Labelling	95
4.3.2	Source Reassignment	95
4.3.3	Experiments	97
4.3.4	Conclusions	97
II	Spatial Sound	99
5	Spatial Sound Systems	101
5.1	Introduction	103
5.2	Spatial Hearing	105
5.2.1	Interaural Differences	105
5.2.2	Spectral Cues	106
5.2.3	Distance and Dynamic Cues	106
5.2.4	The HRTF	107
5.3	Sound Reproduction Systems	108
5.3.1	Stereophony	108
5.3.2	Surround Systems	109
5.3.3	Vector-Based Amplitude Panning	109
5.3.4	Sound Field Rendering	111
5.4	Wave Field Synthesis	111

5.4.1	Kirchhoff-Helmholtz and Rayleigh Integrals	112
5.4.2	Derivation of the Driving Signal Function	116
5.4.3	Potential and Constrains of WFS	118
5.4.4	Special Properties of WFS	120
5.5	Audio Up-Mixing and New Coding Schemes	121
5.5.1	Audio Up-Mixing	121
5.5.2	Spatial Audio Object Coding	122
5.6	Conclusion	124
6	Sound Scene Resynthesis	125
6.1	Introduction	127
6.2	Resynthesis of Sound Scenes	128
6.2.1	Stereo to Wave Field Synthesis Up-Mixing	129
6.3	Sound Scene Evaluation	130
6.3.1	Perceptual Attributes of Spatial Sound	130
6.4	Experiments	133
6.4.1	Stage 1: Stereo Mixing of Reference Sound Scenes	134
6.4.2	Stage 2: Source Separation and Objective Evaluation	134
6.4.3	Stage 3: Resynthesis and Subjective Evaluation	134
6.5	Results	138
6.5.1	Locatedness	138
6.5.2	Localization Accuracy	139
6.5.3	Source Widthness	139
6.5.4	Ensemble Aperture	141
6.5.5	Sound/Timbral Quality	141
6.5.6	Significance analysis with ANOVA	142
6.5.7	Joint Analysis	144
6.5.8	Objective vs. Subjective Measures Discussion	146
6.6	Conclusions	146
7	Localization and Enhanced Reproduction	149
7.1	Introduction	151
7.2	Multiple Source Localization	152
7.2.1	Laplacian Mixture Model	153
7.2.2	Experiments	154
7.3	Stereo Enhancement With Selective Amplitude Panning	156
7.3.1	Time-Frequency Selective Panning	157
7.4	Enhanced Reproduction using Binaural Synthesis	158
7.4.1	System Geometry	158
7.4.2	Full-Azimuth DOA Map	160
7.4.3	Binaural Synthesis	160
7.5	Conclusion	161
8	Conclusions	163
8.1	Summary	165
8.2	Contribution to Knowledge	166
8.3	Further Work	169
8.4	List of Publications	170
	Bibliography	173