



UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

University of Wollongong Thesis Collection
1954-2016

University of Wollongong Thesis Collections

2006

3D-audio object oriented coding

Guillaume Potard
University of Wollongong

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author.

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Recommended Citation

Potard, Guillaume, 3D-audio object oriented coding, PhD thesis, School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, 2006. <http://ro.uow.edu.au/theses/539>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

NOTE

This online version of the thesis may have different page formatting and pagination from the paper copy held in the University of Wollongong Library.

UNIVERSITY OF WOLLONGONG

COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

3D-AUDIO OBJECT ORIENTED CODING

By
Guillaume Potard

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
UNIVERSITY OF WOLLONGONG
NORTHFIELDS AVE
WOLLONGONG NSW 2522
AUSTRALIA
SEPTEMBER 2006

© Copyright by Guillaume Potard, 2006

UNIVERSITY OF WOLLONGONG
DEPARTMENT OF
SCHOOL OF ELECTRICAL, COMPUTER AND
TELECOMMUNICATIONS ENGINEERING

The undersigned hereby certify that they have read and recommend to the Faculty of Faculty of Informatics for acceptance a thesis entitled “**3D-audio object oriented coding**” by **Guillaume Potard** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Dated: September 2006

External Examiner: _____

Research Supervisor: _____
Ian Burnett

Examining Committee: _____
Peter Svensson

Stephen Barrass

UNIVERSITY OF WOLLONGONG

Date: **September 2006**

Author: **Guillaume Potard**

Title: **3D-audio object oriented coding**

Department: **School of Electrical, Computer and
Telecommunications Engineering**

Degree: **Ph.D.** Convocation: **October** Year: **2006**

Permission is herewith granted to University of Wollongong to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

Signature of Author

THE AUTHOR RESERVES OTHER PUBLICATION RIGHTS, AND NEITHER THE THESIS NOR EXTENSIVE EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT THE AUTHOR'S WRITTEN PERMISSION.

THE AUTHOR ATTESTS THAT PERMISSION HAS BEEN OBTAINED FOR THE USE OF ANY COPYRIGHTED MATERIAL APPEARING IN THIS THESIS (OTHER THAN BRIEF EXCERPTS REQUIRING ONLY PROPER ACKNOWLEDGEMENT IN SCHOLARLY WRITING) AND THAT ALL SUCH USE IS CLEARLY ACKNOWLEDGED.

I would to thank my supervisor, Ian Burnett, who guided me through the years of the PhD with useful advice and encouragement.

I then dedicate this thesis to my partner, Megan Sproats, for all her support, patience and love.

Thank you to my parents for helping me to study in Australia and a big hello to my brother.

Table of Contents

Table of Contents	v
List of Tables	xi
List of Figures	xii
1 Introduction	1
1.1 3D audio object oriented coding and rendering	1
1.2 Thesis Outline	4
1.3 Contributions	6
1.4 Publications	8
1.4.1 Conference papers	8
1.4.2 MPEG meeting input papers	9
1.4.3 MPEG meeting output papers	10
2 Encoding and perception of 3D audio	11
2.1 Introduction	11
2.2 Encoding of 3D audio scenes	12
2.3 Channel oriented encoding of 3D audio scenes	15
2.3.1 Binaural recording	15
2.3.2 Multi-channel techniques	16
2.3.3 Ambisonics	19
2.4 Object oriented encoding of 3D audio scenes	22
2.4.1 VRML and X3D	22
2.4.2 MPEG-4	28
2.4.3 Other technologies	36
2.4.4 Summary of 3D audio scene encoding approaches	37
2.5 Spatial auditory perception	38
2.5.1 Localisation	38
2.5.2 Distance perception	42

2.5.3	Other percepts	44
2.5.4	Summary	44
2.6	Introduction to sound source extent perception	45
2.7	Apparent size of a single sound source	45
2.7.1	Effect of pitch on tonal volume	47
2.7.2	Effect of loudness on tonal volume	47
2.7.3	Effect of duration on tonal volume	48
2.7.4	Effect of signal type on tonal volume	50
2.8	Apparent extent of multiple sound sources	50
2.8.1	Overview of the effect	51
2.8.2	Definition of the inter-aural cross-correlation coefficient (IACC)	53
2.8.3	Relationship between the inter sound source correlation coefficients (ISCC) and the IACC	54
2.8.4	Effects of inter sound source coherence	56
2.8.5	Conditions for binaural fusion	58
2.8.6	Multi-dimensionality of sound source extent	64
2.9	Perception of source extent and spaciousness in reverberant environments	66
2.9.1	The precedence effect	66
2.9.2	Spatial Impression	67
2.9.3	Spaciousness	69
2.9.4	Apparent source width	69
2.9.5	Listener envelopment	70
2.9.6	Reverberance	70
2.10	Summary of sound source extent perception	71
2.11	Sound source extent rendering techniques	71
2.11.1	Stereo sound recording techniques	72
2.11.2	Pseudo-stereo processors	73
2.11.3	Ambisonics W Channel boosting	74
2.11.4	Ambisonics O-Format	74
2.11.5	VBAP spread	76
2.12	Rendering of sound source extent using decorrelated point sources	77
2.12.1	Preliminary observations on natural sound sources	77
2.12.2	General principle	78
2.12.3	1, 2 or 3D source extent	78
2.12.4	Extension and evaluation of the decorrelated point source method	79
2.12.5	Obtaining decorrelated signals	79
2.13	Signal decorrelation techniques	80
2.13.1	Time delay	81
2.13.2	Fixed FIR all-pass filters	82
2.13.3	Fixed IIR all-pass filters	86

2.13.4	Feedback Delay Networks	87
2.13.5	Remarks on fixed decorrelation	87
2.13.6	Dynamic decorrelation	88
2.13.7	Frequency varying decorrelation	89
2.13.8	Time varying decorrelation	90
2.13.9	Other decorrelation techniques	91
2.13.10	Summary of source extent rendering techniques	92
2.14	General summary	92
3	Novel object-oriented approach for describing 3D audio scenes using XML	93
3.1	Introduction	93
3.2	XML3DAUDIO: A new 3D audio scene description scheme	95
3.2.1	Design philosophy and aims of XML3DAUDIO	95
3.2.2	3D audio scene description areas	98
3.3	The scene orchestra and score approach	100
3.3.1	Scene orchestra: content description	100
3.3.2	Scene score: initialisation, timing, composition and hierarchy description	100
3.3.3	Benefits of the scene orchestra and score approach	102
3.3.4	Format of the scene orchestra	102
3.3.5	Format of the scene score	105
3.3.6	List of scene orchestra objects	111
3.3.7	3D audio scene example	126
3.4	Evaluation of the novel scheme	130
3.4.1	Feature comparison with VRML and MPEG-4	130
3.4.2	Simplification of scene description by the novel scheme	131
3.4.3	Description of hybrid 3D audio scenes	138
3.5	Use of the proposed scheme as a meta-data annotation scheme for 3D audio content	140
3.5.1	Introduction	140
3.6	Summary	144
4	Perception of sound source extent and shape	145
4.1	Introduction	145
4.2	Overview of the experiments	147
4.3	Experiment 1: Perception of one-dimensional horizontal sound source extent	149
4.3.1	Aims	149
4.3.2	Apparatus	149

4.3.3	Stimuli	150
4.3.4	Procedure	151
4.3.5	Results	154
4.3.6	Analysis of Results	160
4.3.7	Discussion	174
4.4	Experiment 2: perception of horizontal, vertical and 2D sound source extent	178
4.4.1	Aims	178
4.4.2	Apparatus	179
4.4.3	Stimuli	181
4.4.4	Procedure	182
4.4.5	Results	183
4.4.6	Discussion	185
4.5	Experiment 3: perception of sound source shape using real decorrelated sound sources	187
4.5.1	Aims	187
4.5.2	Apparatus	188
4.5.3	Stimuli	191
4.5.4	Procedure	192
4.5.5	Results	192
4.5.6	Analysis of Results	196
4.5.7	Discussion	199
4.6	Experiment 4: perception of sound source shape using virtual decorrelated sound sources	202
4.6.1	Aims	202
4.6.2	Apparatus	203
4.6.3	Stimuli	203
4.6.4	Procedure	206
4.6.5	Results	206
4.6.6	Analysis of results	209
4.6.7	Discussion	211
4.7	Experiment 5: improvement in 3D audio scene realism by using extended sound sources	213
4.7.1	Aims	213
4.7.2	Apparatus	214
4.7.3	Stimuli	214
4.7.4	Procedure	215
4.7.5	Results	215
4.7.6	Discussion	216
4.8	Experiment 6: perceptual effects of dynamic decorrelation	216

4.8.1	Aims	216
4.8.2	Apparatus	216
4.8.3	Stimuli	216
4.8.4	Procedure	217
4.8.5	Results	217
4.8.6	Discussion	218
4.9	Experiment 7: perceptual effects of time-varying decorrelation	219
4.9.1	Aims	219
4.9.2	Apparatus	219
4.9.3	Stimuli	219
4.9.4	Procedure	220
4.9.5	Results	221
4.9.6	Discussion	222
4.10	Implementation of sound source extent description capabilities in MPEG-4 AudioBIFS	223
4.11	Summary	226
5	Implementation of an object oriented 3D audio scene renderer	229
5.1	Introduction	229
5.2	CHES system overview	231
5.2.1	Speaker vs headphone 3D audio rendering	232
5.2.2	CHES speaker array	233
5.2.3	Hardware	234
5.3	Digital signal processing layer	235
5.3.1	Selection of the rendering platform	236
5.3.2	3D audio signal processing overview	237
5.4	Description of 3D audio processing tasks used in CHES	243
5.4.1	Spatialisation	243
5.4.2	Implementation of 4th order Ambisonics spatialisation in CHES	257
5.4.3	Sound source distance rendering	260
5.4.4	Sound source extent rendering	263
5.4.5	Propagation delays and Doppler effect	264
5.4.6	Sound source occlusion	267
5.4.7	Early reflections calculation	269
5.4.8	Late reverberation	272
5.5	Scene manager	274
5.6	Evaluation	281
5.6.1	3D audio rendering quality	281
5.6.2	System structure	285
5.7	Practical uses of CHES	287

5.8	Summary	290
6	Conclusions and further work	291
6.1	3D audio scene description	291
6.2	Sound source extent and shape	292
6.3	3D audio rendering	294
6.4	General conclusion	295
	Bibliography	297
7	Appendix A	322
7.1	Measurements of inter-signal correlation	322
7.2	Matlab code for IIR decorrelation filter	323
7.3	Matlab code for dynamic decorrelation filter	324
8	Appendix B	326
8.1	List of DSP layer commands	326
8.1.1	Sound source control	326
8.1.2	Reflective surface control	327
8.1.3	Room reverberation control	327

List of Tables

2.1	List of AudioBIFS nodes	34
2.2	List of Advanced AudioBIFS nodes	36
3.1	List of orchestra objects	105
3.2	Examples of two initialisation score lines	106
3.3	Examples of performance score lines	108
4.1	Percentages of preference in terms of naturalness between fixed and dynamic decorrelation	218
4.2	Average listening fatigue caused by fixed, dynamic and no decorrelation (1: no fatigue, 5: extreme fatigue)	218
4.3	Average frequencies of the rate of change of the IACC at which subjects could no more perceive a change in source extent	221
5.1	Spherical harmonics encoding equations up to Ambisonics order 4, (Encoding equations up to order 3 obtained from [Dan00])	249
5.2	Coordinates of the CHESSE speakers	256

List of Figures

2.1	Transmission of 3D audio content using the channel oriented approach	12
2.2	Transmission of 3D audio content using the object-oriented approach	13
2.3	Dummy head microphone example for recording 3D audio scenes bin-aurally (Neumann KU100 model)	15
2.4	5.1 Surround speaker positioning as defined by the ITU BS.775-1 recommendation	17
2.5	Schematic view of the B-format W,X,Y and Z channel directivity patterns	19
2.6	Overview of the Ambisonics encoding/decoding approach	20
2.7	Tetrahedral configuration of capsules inside the Soundfield microphone	21
2.8	Overview of the VRML server/client architecture	24
2.9	Schematic view of a scene graph	25
2.10	Semantics of the VRML sound nodes	27
2.11	VRML sound source ellipsoidal directivity model with only four parameters	28
2.12	Illustration of an animation circuit in VRML	29
2.13	Standardised MPEG-4 system layers between raw bitstream and renderer	30
2.14	Example of BIFS scene containing video, audio, text and a graphical user interface	31
2.15	Illustration of BIFS-Commands and BIFS-Anim streams animating modifying the state of the scene graph in a timely manner	33
2.16	Illustration of the AudioBIFS input, composition and output nodes	35

2.17 a) Use of Inter-aural time differences (ITD) at low frequencies, b) Use of Inter-aural level differences (ILD) at high frequencies	39
2.18 Summing localisation results in the localisation of a phantom sound source in the presence of multiple coherent sound sources	42
2.19 In reverberant conditions, localisation of the main sound source is preserved thanks to the precedence effect which inhibits the perception of reflections which reach the listener after the direct sound	43
2.20 Illustration of the difference between the physical size of a sound source and its perceived tonal volume	46
2.21 Decrease in tonal volume for an increase in frequency of a pure sine tone, at three stimulus durations (reproduced with permission from [PB82])	48
2.22 Increase in perceived tonal volume with increase in stimuli loudness and duration (reproduced with permission from [PB82])	49
2.23 Illustration of the extent of multiple sound sources: a) Coherent sound sources result in a narrow source extent at the centre of gravity, b) Incoherent sound sources result in a broad extent	53
2.24 Relationship between the inter-source cross-correlation coefficients (ISCC) and the interaural cross-correlation coefficient(IACC)	55
2.25 Effect of the inter-aural cross-correlation coefficient (IACC) on the apparent image width of white noise presented on headphones	57
2.26 Effect of inter-channel correlation on perceived spatial extent	58
2.27 Effects of angular separation between two uncorrelated sound sources on apparent source extent and binaural fusion: a) Perception of a single narrow auditory event, b) Perception of a single broad auditory event, c) Perception of two distinct auditory events	60
2.28 General model of the conditions affecting binaural fusion and apparent extent of multiple sound sources	63
2.29 Example of one-dimensional and multidimensional sound sources	64

2.30	Simplified model of room reverberation	67
2.31	Illustration of the precedence effect (after Kendal [Ken95])	68
2.32	Relationship between ‘Spatial Impression’ and other auditory percepts	68
2.33	Apparent Source Width (ASW) and localisation blur increase with distance in reverberant conditions	70
2.34	Illustration of the MS stereophonic microphone recording to capture and control image width	73
2.35	Inwards and outwards equivalence between B-format and O-format .	75
2.36	Sampling of the directivity pattern and shape extent of a sound source with a microphone array prior to O-format conversion	76
2.37	Decomposition of a vibrating panel source into several point sound source	77
2.38	Creation of 1D, 2D and 3D broad sound sources using the decorrelated point source method	78
2.39	Capture and reproduction of the extent of a natural sound source via a microphone array	80
2.40	Capture and reproduction of the extent of a natural sound source via a single microphone and a decorrelation filterbank	80
2.41	Obtaining decorrelated signals by delaying an input signal	82
2.42	Decorrelation filterbank to create several uncorrelated replicas of a monaural signal	83
2.43	Frequency and phase response of an all-pass FIR decorrelation filter .	84
2.44	Impulse response of an all-pass FIR decorrelation filter	84
2.45	Obtaining an all-pass FIR decorrelation filter via artificial magnitude and phase response construction and inverse Fast Fourier Transform .	86
2.46	Architecture of an order 3 feedback delay network	88
2.47	Principle of a sub-band decorrelator	90
2.48	Principle of a time-varying decorrelator	91
3.1	Illustration of the three description categories to describe 3D audio scenes	99

3.2	Overview of the orchestra and score approach	101
3.3	Format of the scene orchestra	104
3.4	Scene score format	106
3.5	Formats of the lines of initialisation and performance score	107
3.6	List of scene score commands	110
3.7	Semantics of the <i>Listener</i> object	112
3.8	Example of sound source directivity described at two frequencies . . .	113
3.9	Semantics of the <i>Source</i> object	115
3.10	Semantics of the <i>Surface</i> object	117
3.11	Semantics of the <i>Medium</i> object	119
3.12	Parameters of the <i>Room</i> object	120
3.13	Illustration of car macro-object	121
3.14	Macro-object definition schema	122
3.15	Semantics of the <i>macro-object</i> object used to import complex objects in the scene orchestra	123
3.16	Semantics of the <i>recorded scene</i> object	125
3.17	Semantics of the <i>definition</i> objects	126
3.18	Comparison of 3D audio scene description capabilities between VRML, MPEG-4 AudioBIFS and the novel scheme	131
3.19	Playing times and animation of the example 3D audio scene	132
3.20	Scene orchestra and score description of the 3D audio scene example .	133
3.21	Scene graph description of the 3D audio scene example	134
3.22	Illustration of the hybrid 3D audio scene rendering process	139
3.23	Illustration of XML meta-data generation from the 3D audio scene description	143
3.24	Generation of 3D audio meta-data at the authoring stage	143
4.1	Seven-speaker horizontal array apparatus used in the experiment study- ing the perception of horizontal sound source extent	150

4.2	Construction of 21 horizontally extended sound source stimuli using three different densities of decorrelated point sources	152
4.3	Answer sheet for drawing the perceived horizontal extents of the presented stimuli	153
4.4	Mean perceived extent of 0 degree extended stimuli for four types of signals	154
4.5	Mean perceived extent of 10 degree extended stimuli for four types of signals at two different point source densities	155
4.6	Mean perceived extent of 30 degree extended stimuli for four types of signals at three different point source densities	155
4.7	Mean perceived extent of 60 degree extended stimuli for four types of signals at three different point source densities	156
4.8	Mean perceived extent of 90 degree extended stimuli for four types of signals at three different point source densities	156
4.9	Mean perceived extent of 120 degree extended stimuli for four types of signals at three different point source densities	157
4.10	Mean perceived extent of 150 degree extended stimuli for four types of signals at three different point source densities	157
4.11	Mean perceived extent of 180 degree extended stimuli for four types of signals at three different point source densities	158
4.12	Mean perceived horizontal extent of the 21 stimuli across the four signal types	159
4.13	Mean Error and 95% confidence intervals between perceived and actual source width for 3 point source density	161
4.14	Mean Error and 95% confidence intervals between perceived and actual source width for one sound source per 10 degree density	162
4.15	Mean Error and 95% confidence intervals between perceived and actual source width for one sound source per 30 degree density	163

4.16	Mean Error and 95% confidence intervals between perceived and actual source width at three sound source densities	164
4.17	Grand mean error and 95% confidence intervals between perceived and actual source width at three sound source densities	165
4.18	Grand mean error and 95% confidence intervals between perceived and actual source width for the two stimulus signal types	165
4.19	Grand mean error and 95% confidence intervals between perceived and actual source width for the two stimulus levels	166
4.20	3-Factor ANOVA: F-ratios and confidence interval for 3 point source density	168
4.21	3-Factor ANOVA: F-ratios and confidence intervals for 1 source per 10 degree density	169
4.22	3-Factor ANOVA: F-ratios and confidence intervals for 1 source per 30 degree density	170
4.23	4-Factor ANOVA: F-ratios and confidence intervals	171
4.24	Mean of stimuli perceived as one sound source	172
4.25	Percentage of answers where stimuli were perceived as single sound sources	173
4.26	Geometry of the 16-speaker array apparatus	180
4.27	Position of the subjects in relation to the apparatus	181
4.28	Answer sheet for the 2D sound source extent experiment	183
4.29	Distribution of perceived source extents and mean percentages of on-target answers for 1D and 2D sound sources presented on a three-dimensional auditory display	184
4.30	Position of the speaker array in relation to the subjects	188
4.31	Diagram of the speaker array	189
4.32	Coordinates of the decorrelated point sources/speakers	189

4.33	Apparatus of the sound source shape perception experiment with real decorrelated sound sources. From left to right: at Thomson (Germany), University of Wollongong and ETRI (Korea)	190
4.34	Geometry of the six sound source shapes used in the experiment . . .	192
4.35	Percentages of correct sound source shape identifications (not including shape ‘A’)	193
4.36	Confusion matrices of sound source shape identification for the four signal types (frontal stimulus presentation)	194
4.37	Confusion matrices of sound source shape identification for the four signal types (rear stimulus presentation)	195
4.38	Mean percentage and 95% confidence interval of correct shape identification across four signal types, stimuli presented behind subjects . .	196
4.39	Mean percentage and 95% confidence interval of correct shape identification across four signal types, stimuli presented in front of subjects	197
4.40	Mean percentage and 95% confidence interval of correct shape identification in function of sound source shape type, stimuli presented behind subjects (results averaged across the four signal types)	197
4.41	Mean percentage and 95% confidence interval of correct shape identification in function of sound source shape type, stimuli presented in front of subjects (results averaged across the four signal types)	198
4.42	Grand mean percentage and 95% confidence interval of correct shape identification stimuli presented in the back and in front of subjects . .	198
4.43	3-Factor ANOVA: F-ratios and confidence intervals	199
4.44	Placement of subjects at the centre of the speaker cube apparatus . .	204
4.45	Geometry of the five sound source shapes	205
4.46	Coordinates of the point sources used to form the sound source shapes	206
4.47	Confusion matrix of shape identifications (shapes created with decorrelated virtual sound sources)	207

4.48	Confusion matrix of shape identifications (shapes created with correlated virtual sound sources)	208
4.49	Mean percentage and 95% confidence interval of correct shape identification for decorrelated and correlated point sound sources	209
4.50	Mean percentage and 95% confidence interval of correct shape identification for the five sound source shapes for decorrelated point sound sources)	210
4.51	Mean percentage and 95% confidence interval of correct shape identification for the five sound source shapes for correlated point sound sources)	210
4.52	2-Factor ANOVA: F-ratios and confidence intervals	211
4.53	Percentages of time where 3D audio scenes that used extended sound sources were subjectively preferred (for speaker and headphone stimulus presentation)	215
4.54	Mean rate of change of IACC and 95% confidence interval at which subject perceived no more change in sound image width	221
4.55	Different sound source shape types definable in the field of the WideSound node	224
4.56	Semantics of the new <i>WideSound</i> AudioBIFS node to represent sound sources with apparent extents in MPEG-4 AudioBIFS scenes	225
4.57	MPEG-4 AudioBIFS 3D audio scene example containing four <i>WideSound</i> nodes	226
5.1	Overview of the client-server structure of the CHESS system and the functions and technologies of the different system parts	232
5.2	The configurable speaker array of the CHESS system	234
5.3	Graphical user interface of the DSP layer	235
5.4	Overview of the signal processing chain in CHESS for the calculation of direct sound, reflections and reverberation for one sound source	239
5.5	Signal processing chain for the direct signal path	241

5.6	Spherical coordinate system used in the CHESS system	245
5.7	Illustration of the Higher Order Ambisonics encoding operation . . .	248
5.8	Illustration of the forming of an audio scene by adding n Ambisonics signals from k encoded sound sources	250
5.9	Diagram of the Ambisonics decoding process via a decoding matrix D	253
5.10	Diagram of the icosahedron polyhedra used to place speakers (upper hemisphere used only)	254
5.11	Numbering and placement of speakers in CHESS (top-down view) . .	255
5.12	Patch of the new Max/Msp object for performing 4th order Ambisonics encoding in CHESS	258
5.13	Patch of the new Max/Msp patch for performing 4th order Ambisonics decoding in CHESS	259
5.14	Interface of the 4th Order Ambisonics spatialisation plugin in Protools	260
5.15	Illustration of distance control and the minimum source distance in CHESS	262
5.16	Illustration of simple horizontal source extent rendering in CHESS . .	264
5.17	New decorrelation object for Max/Msp	265
5.18	Diagram of a variable delay line to implement delay and Doppler effects	266
5.19	Detection of sound source occlusion by a surface object	268
5.20	Algorithm for sound source occlusion detection in CHESS	268
5.21	Illustration of the image model algorithm principle	269
5.22	Diagram of the first order image model algorithm used in CHESS . .	271
5.23	Perceptual control of room reverberation in CHESS	273
5.24	Overview of the scene manager structure	275
5.25	Diagram of 3D audio scene rendering from an XML scene description in CHESS	276
5.26	Illustration of the score modifying the current state of the orchestra in the scene renderer memory	278
5.27	Graphical interface of the Java3D scene manager	280

5.28	Listening to the mind listening promotion	288
5.29	Picture of the outdoor CHESS dome at Sonic Connections 2004 . . .	289

ABSTRACT

This thesis first presents a novel object-oriented scheme which provides for extensive description of time-varying 3D audio scenes using XML. The scheme, named XML3DAUDIO, provides a new format for encoding and describing 3D audio scenes in an object oriented manner. Its creation was motivated by the fact that other 3D audio scene description formats are either too simplistic (VRML) and lacking in realism, or are too complex (MPEG-4 Advanced AudioBIFS) and, as a result, have not yet been fully implemented in available decoders and scene authoring tools. This thesis shows that the scene graph model, used by VRML and MPEG-4 AudioBIFS, leads to complex and inefficient 3D audio scene descriptions. This complexity is a result of the aggregation, in the scene graph model, of the scene content data and the scene temporal data. The resulting 3D audio scene descriptions, are in turn, difficult to re-author and significantly increase the complexity of 3D audio scene renderers. In contrast, XML3DAUDIO follows a new scene orchestra and score approach which allows the separation of the scene content data from the scene temporal data; this simplifies 3D audio scene descriptions and allows simpler 3D audio scene renderer implementations. In addition, the separation of the temporal and content data permits easier modification and re-authoring of 3D audio scenes. It is shown that XML3DAUDIO can be used as a new format for 3D audio scene rendering or can alternatively be used as a meta-data scheme for annotating 3D audio content.

Rendering and perception of the apparent extent of sound sources in 3D audio displays is then considered. Although perceptually important, the extent of sound sources is one the least studied auditory percepts and is often neglected in 3D audio displays. This research aims to improve the realism of rendered 3D audio scenes by reproducing the multidimensional extent exhibited by some natural sound sources (eg a beach front, a swarm of insects, wind blowing in trees etc). Usually, such broad

sound sources are treated as point sound sources in 3D audio displays, resulting in unrealistic rendered 3D audio scenes. A technique is introduced whereby, using several uncorrelated sound sources, the apparent extent of a sound source can be controlled in arbitrary ways. A new hypothesis is presented suggesting that, by placing uncorrelated sound sources in particular patterns, sound sources with apparent shapes can be obtained. This hypothesis and the perception of vertical and horizontal sound source extent are then evaluated in several psychoacoustic experiments. Results showed that, using this technique, subjects could perceive the horizontal extent of sound sources with high precision, differentiate horizontally from vertically extended sound sources and could identify the apparent shapes of sound sources above statistical chance. In the latter case, however, the results show identification less than 50 % of the time, and then only when noise signals were used. Some of these psychoacoustic experiments were carried out for the MPEG standardisation body with a view to adding sound source extent description capabilities to the MPEG-4 AudioBIFS standard; the resulting modifications have become part of the new capabilities in version 3 of AudioBIFS.

Lastly, this thesis presents the implementation of a novel real-time 3D audio rendering system known as CHESS (Configurable Hemispheric Environment for Spatialised Sound). Using a new signal processing architecture and a novel 16-speaker array, CHESS demonstrates the viability of rendering 3D audio scenes described with the XML3DAUDIO scheme. CHESS implements all 3D audio signal processing tasks required to render a 3D audio scene from its textual description; the definition of these techniques and the architecture of CHESS is extensible and can thus be used as a basis model for the implementation of future object oriented 3D audio rendering systems.

Thus, overall, this thesis presents contributions in three interwoven domains of 3D audio: 3D audio scene description, spatial psychoacoustics and 3D audio scene rendering.