Goldsmiths
UNIVERSITY OF LONDON

A Perceptual Approach to Audio-Visual Instrument Design, Composition and Performance

Adriana de Almeida Portela Viana de Sá

PhD Thesis in Arts and Computational Technologies, Department of Computing

Supervisors: Atau Tanaka (1st) and Mick Grierson (2nd)

I declare that this thesis was composed by myself and that the work contained herein is my own, except where explicitly stated otherwise in the text.

Adriana Sa

ABSTRACT

The thesis presents a perceptual approach to audio-visual instrument design, composition and performance. The approach informs practical work as well as a parametric visualisation model, which can be used to analyse sensory dominance, sonic expression and spatial presence in any audio-visual performance language.

The practical work intends for the image to function as a stage scene, which reacts to the music and allows for attention to focus on the relation between the sounds themselves. This is challenging, because usually vision dominates over audition. To clarify the problem, the thesis extrapolates from audio-visual theory, psychology, neuroscience, interaction design and musicology. The investigations lead to three creative principles, which inform the design of an instrument that combines a custom zither and audio-visual 3D software. The instrument uses disparities between the acoustic and digital outputs so as to explore those creative principles: a) to threshold the performer's control over the instrument and the instrument's unpredictability, in ways that convey musical expression; b) to facilitate perceptual simplification of visual dynamics; c) to create an audio-visual relationship that produces a sense of causation, and simultaneously confounds the cause and effect relationships. This latter principle is demonstrated with a study on audio-visual mapping and perception, whose conclusions are equally applicable to the audio-visual relationship in space. Yet importantly, my creative decisions are not driven by demonstrative aims. Regarding the visual dynamics, the initial creative work assures perceptual simplification, but the final work exposes a gray area that respects to how the audience's attention might change over time.

In any case, the parametric visualisation model can reveal how any audio-visual performance work might converge or diverge from these three creative principles. It combines parameters for interaction, sonic & visual dynamics, audio-visual relationship, physical performance setup and semantics. The parameters for dynamics and semantics reflect how stimuli inform attention at a particular timescale. The thesis uses the model to analyse a set of audio-visual performance languages, to represent my solo performance work from a creative perspective, and to visualise the work's versatility in collaboration with other musicians.

Keywords:

NIME, audio-visual performance, music, 3D environments, perception, attention

ACKNOWLEDGEMENTS

I would like to thank my supervisors Atau Tanaka and Mick Grierson for their remarks and advices. John Klima was an indispensable collaborator in my practical work. Baptiste Caramiaux' advice was important in the design of a controlled experiment. I would also like to thank the participants in that experiment. Equal thanks to Joanne Cannon, Helena Espvall, Tó Trips and Aki Onda, for their participation in case studies. Also, thanks to all who helped in the preparation of articles and papers, because their comments led me to clarify and strengthen the argument of this thesis: Steve Gibson, Stefan Mueller Arisona, Baptiste Caramiaux, Atau Tanaka, Mick Grierson, John Klima, Alex McLean, Brigitta Zics, Rajele Jain, Alan Klima, Adam Parkinson, Ana Portela, and all anonymous reviewers. Finally, thanks to Freida Abtan and Joseph Heyde for their excellent *viva voce* examination, which also pointed to further research vectors.

This research was funded by the FCT - Foundation of Science and Technology and the European Union, through the operational program POPH/FSE (ref. SFRH/BD/68632/2010).

TABLE OF CONTENTS

List of Figures 11
List of Tables 13
List of Photographs 13
Publications and Related Activities 16
Documentation and Support Material 20

Chapter 1: Introduction

1.1 An Artistic Position 23
  1.1.1 A Personal Context 23
  1.1.2 A Personal Audio-Visual Instrument 24
1.2 Research Aims and Questions 25
1.3 Starting Points and Contributions 28
  1.3.1 The Advantage of a Parametric Visualisation Model 28
  1.3.2 Sonic Expression: Creative Motivations & Theoretical Contributions 29
  1.3.3 Sensory Dominance: Creative Motivations & Theoretical Contributions 30
  1.3.4 Spatial Presence: Creative Motivations & Theoretical Contribution 31
  1.3.5 Practical Work 31
1.4 Thesis Structure 32

Chapter 2: Literature Review and Research Methods

2.1 Chapter Introduction 35
2.2 Instrument and Spatial Presence 35
  2.2.1 Instrument as Open-Ended System 37
  2.2.2 Instrument as Mediation 37
  2.2.3 Spatial Presence and Performative Arena 39
2.3 Audio-Visual Relationships and Sensory Dominance 41
  2.3.1 Approaches to Audio-Visual Performance and Animation 41
  2.3.2 Music with Digital 3D Environments 45
  2.3.3 The Artistic Debate on Sensory Dominance 48

2.4 Perception, Structure and Complexity 51
2.4.1 Perception 51
2.4.2 Multisensory Integration 54
2.4.3 Attention 56
2.4.4 Audio-Visual Objecthood 58
2.4.5 Intensity 60
2.5 Sonic Expression, Interaction and Flow 61
2.5.1 Interaction 61
2.5.2 Sonic Expression: "Chance", Intensity, Effort and Timing 62
2.5.3 Flow 63
2.5.4 Flow in music 65
2.6 Discussion and Methods 69

Chapter 3: Creative Principles

3.1 Chapter Introduction 75
3.2 The Experience of Complexity: Intention vs. Intentionality 76
3.3 From Instrument Design to Sonic Expression 78
3.3.1 The Two-Folded Interaction with Musical Instruments 78
3.3.2 Musical “Grid”, Effort and Dynamic Deviation 81
3.4 Relating Intensity, Attention, Continuities and Discontinuities 83
3.4.1 (Re)defining Intensity in Cross-Sensory Terms 83
3.4.2 Types of Continuities and Discontinuities 86
3.4.3 Using Continuities and Discontinuities 89
3.5 Synchrony, Perceptual Binding and Fungible Audio-Visual Relationships 92
3.6 Study about Audio-Visual Mapping and Perceptual Experience 97
3.6.1 Problem 97
3.6.2 Hypothesis 98
3.6.3 Stimuli 99
3.6.4 Procedure 101
3.6.5 Results 101
3.6.6 Analysis and Discussion 104
3.6.7 Study conclusions 106
3.7 Three Creative Principles for Audio-Visual Instrument Design and Composition ... 107
3.8 Chapter Summary and Discussion 108

Chapter 4: A Parametric Model on Audio-Visual Performance

4.1 Chapter Introduction 112
4.2 Useful Parametric Visualisation Models 112
4.2.1 A Model on Human-Computer Interaction 112
4.2.2 A Model on the Epistemic Dimension of Human-Computer Interaction .. 114
4.2.3 A Model on Electronic Music Performance as Audio-Visual Expression . 116
4.3 A Parametric Model on Sensory Dominance, Sonic Expression and Spatial Presence 119
4.3.1 Learning from Problems: General Considerations about Parameters and Graphical Aspect 119
4.3.2 Parameterising Sonic Expression & the Principle of Sonic Complexity 121
4.3.3 Parameterising Sensory Dominance & the Principles of Visual Continuity and Audio-Visual Fungibility 123
4.3.4 Parameterising Spatial Presence 125
4.3.5 Parameter Visualisation 128
4.4 Analysing Different Audio-Visual Performance Languages 131
4.4.1 Performance by Steina Vasulka 131
4.4.2 Performance by Ryoji Ikeda 135
4.4.3 Performance by Thor Magnusson 137
4.5 Chapter Discussion 140

Chapter 5: Developing an Audio-Visual Instrument and Performance Language

5.1 Chapter Introduction 144
5.2 Instrument Version #1 145
5.2.1 Modifying Video Game Software in Combination with a Zither 146
5.2.2 Sonic Complexity 149
5.2.3 Visual Continuity 152
5.2.4 Fungible Audio-Visual Mapping 153
5.2.5 Performative Arena & the Fungible Audio-Visual Relationship in Space . 154
5.2.6 Perceptual Integration of the Sonic and the Visual Motion 155
5.2.7 Notes: Towards an Expansion of Creative Strategies 156

5.3 Sound Organisation: the Arpeggio-Detuning 157
5.3.1 Amplitude and Pitch Analysis 158
5.3.2 The Zither Tuning and the Arpeggio-Detuning 160
5.3.3 The Interfaces: Technical Aspects of Control and Instability 163
5.3.4 Zither Techniques, Digital Sounds and Semantics 165
5.3.5 Musical Forms and Digital Constraints 166
5.4 Extending Creative Strategies to the Image and Audio-Visual Relationship: AG#2 169
5.5 3D World and Visual Motion 171
5.6 Audio-Visual Mappings 180
5.7 Sound Spatialisation 183
5.7.1 Applying Perceptual Cues for Direction 185
5.7.2 Applying Perceptual Cues for Distance 186
5.8 Sound Diffusion Quality 187
5.8.1 Microphone, Compression and Reverb 187
5.8.2 Testing Speakers and Amplifiers 188
5.9 Chapter Summary and Conclusions 190

Chapter 6: Discussion

6.1 Chapter introduction 193
6.2 Discussing Parameters 193
6.2.1 About the Parameterisation of Sonic Expression 193
6.2.2 About the Parameterisation of Sensory Dominance 196
6.2.3 About the Parameterisation of Spatial Presence 199
6.3 Representing the Practical Work 202
6.4 Analysing the Work's Musical Versatility: Four Case Studies 207
6.4.1 Question, Procedure, and a Distinction Between Expressive and Structural Changes in Music 208
6.4.2 Case Study #1: Collaboration with Joanne Cannon 210
6.4.3 Case Study #2: Collaboration with Helena Espvall 214
6.4.4 Case Study #3: Collaboration with Tó Trips 219
6.4.5 Case Study #4: Collaboration with Aki Onda 223
6.4.6 Representing Versatility 227

6.5 Summary of contributions 229
6.5.1 Sensory Dominance: Summary of Contributions 230
6.5.2 Sonic Expression: Summary of Contributions 231
6.5.3 Spatial Presence: Summary of Contributions 233
6.5.4 Practical Usefulness of the Parametric Model 233

Conclusion 235

Bibliography 241

FIGURES, TABLES AND PHOTOGRAPHS

List of Figures

Figure 1: An arguable distinction between the notion of flow in music and gaming, p.68

Figure 2: The Experience of complexity, p. 76

Figure 3: How linear interface behaviours convey intention and non-linear behaviours convey intentionality, p.80

Figure 4: Intensity depends on the event itself, the stimuli panorama, and perceptual resolution, p.84

Figure 5: Articulating the sonic and the visual motion such that the experience is driven through sound organisation, p.91

Figure 6 (from study): Prototypes 1, 2 and 3, which exhibit one sphere; and Prototype 4, which exhibits two spheres, p.99
(http://doc.gold.ac.uk/~map01apv/StudyMappingPrototypes.mp4)

Figure 7 (from study): Relation between Causation and Transparency for each mapping prototype, p.102

Figure 8: Analysis of the study results, p. 106

Figure 9: Parametric model proposed by Birnbaum et al., 2005, p.113

Figure 10: Parametric model proposed by Magnusson, 2010, p.115

Figure 11a: Parametric model proposed by Ciciliani, 2014, p.117

Figure 11b: Second iteration of the model by Ciciliani, with Mojzysz, 2015, p.117

Figure 12: Birnbaum's parameters represented in a Cartesian graphic, rather than a star-like graphic, p. 120

Figure 13: Factors that influence sonic expression, p.121

Figure 14: Factors that influence sensory dominance, p.123

Figure 15: Factors that influence the sense of spatial presence, p.126

Figure 16: Visualisation of the parameters from which we can infer sonic expression, sensory dominance and spatial presence, p.128

Figure 17: Steina Vasulka's "Violin Power" performance at Strange Music for Nam June Paik, 132

Figure 18: Ryoji Ikeda's performance "Superposition", p.135

Figure 19: Thor Magnusson's "Fermata" performance with his Threnoscope system, p.138

Figure 20: Technical diagram of the 1st version of the audio-visual instrument developed in this research, p.147

Figure 21: Screenshots from AG#1 (2012), p.148

Figure 22: The disparities between the direct output sound of the zither and the digitally mediated outputs from AG#1 convey the principles of sonic complexity and audio-visual fungibility, p.149

Figure 23: Graphic score for a performance with the audio-visual instrument (2011), p.152

Figure 24: Two sine waves with different amplitude, but same frequency and wavelength, p.159

Figure 25: Technical diagram - zither with Arpeggio-Detuning software, p.163

Figure 26: Technical diagram of the audio-visual instrument with AG#2, p.170

Figure 27: 3D world structure in AG#2, p.172

Figures 28 and 29: Alternative background images (for 3D worlds #1 and #2), p.173

Figure 30: Sky dome image, p.173

Figures 31 and 32: Alternative environment terrain images (for 3D worlds #1 and #2), p.174

Figures 33 and 34: Alternative foreground terrain images (for 3D worlds #1 and #2), p.174

Figure 35: Image that establishes the surface qualities of the environment terrain and the foreground terrain (refraction of light source). The RGB represents an XYZ vector indicating an angle and intensity of refraction, p.175

Figure 36: Image that establishes the primary coloration and transparency of the water, p.175

Figure 37: Image applied to the foreground terrain, describing discrete light events, such as refractions of light through water (stored in the environment map's alpha channel), p. 175

Figure 38: Screenshots from an early version of the 3D world made for AG#2, p.176

Figures 39 and 40: Screenshots from two recent 3D worlds, #1 and #2 (terrain elevations are being activated by audio detection), p.177

Figure 41: Inverted double-stereo system – a simple way of exploring the principle of audio-visual fungibility with 3D sound, p.184

Figure 42: Area-based sound, p.186

Figure 43: Using Magnusson's model to represent interaction with the first and the final version of the instrument, p.194

Figure 44: Parametric visualisation of a solo performance, p.203

Figure 45: Parametric visualisation of the musical collaboration with Joanne Cannon, p. 212

Figure 46: Ecological fit intrinsic to Joanne Cannon’s’s Serpent, p. 213

Figure 47: Parametric visualisation of an hypothetic audio-visual performance with Joanne Cannon, p. 213

Figure 48: Parametric visualisation of the musical collaboration with Helena Espvall, p. 218

Figure 49: Ecological fit intrinsic to Helena Espvall's cello playing, p. 219

Figure 50: Parametric visualisation of an hypothetic audio-visual performance with Helena Espvall, p. 219

Figure 51: Parametric visualisation of the musical collaboration with Tó Trips, p.222

Figure 52: Ecological fit intrinsic to Tó Trips' guitar playing, p.222

Figure 53: Parametric visualisation of an hypothetic audio-visual performance with Tó Trips, p.222

Figure 54: Parametric visualisation of the musical collaboration with Aki Onda, p.225

Figure 55: Ecological fit intrinsic to Aki Onda playing his electronic devices, p.226

Figure 56: Parametric visualisation of an hypothetic audio-visual performance with Aki Onda, p.226

Figure 57: Parametric visualisation of the work’s versatility in collaboration with other musicians, p.228

List of Tables

Table 1: Gestaltist principles of perceptual organisation, p.52

Table 2: Perceptual integration of the sonic and the visual motion, p.92

Table 3: Average ratings for Causation and Transparency, p.102

Table 4: Combinations of sample banks and zither techniques, p.166

Table 5: 3D world structure and visual dynamics in AG#2, p.172

Table 6: The fungible audio-visual mapping in AG#2, p.181

List of Photographs

Photo 1: Performance with the 1st version of the audio-visual instrument at Sho-Zyg (“Immerge”), St. James's Hatcham Church, London, 2012, p. 146

Photo 2: Zither and piezo microphone attached to its body, p.148

Photo 3: Performance with the 1st version of the audio-visual instrument at In-Sonora, Madrid, 2011, p.154

Photos 4 and 5: Setup including zither, piezo microphone, compression pedal & iPad with the Arpeggio-Detuning software, p.158

Photo 6: Playing the zither with the Arpeggio-Detuning software, 195 Lisbon Studio, 2013, p.164

Photos 7-12: Performance with the final version of the instrument at ScratchBuilt Studios, Lisbon, 2014, p.178

Photo 13: Piezo microphone, p.187

Photo 14: Compressor, p.188

Photo 15: Reverb, p.188

Photo 16: Amplifier and cabinet, p.189

Photo 17: Solo performance at Amersham Arms, London, 2013, p.206

Photo 18: Joanne Cannon playing the Serpent, p.210

Photo 19: Performance with Joanne Cannon at Trem Azul, Lisbon, 2013, p. 211

Photo 20: Recording with Helena Espvall at ScratchBuilt Studio, Lisbon, 2014, p.216

Photo 21: Performance with Tó Trips at Trem Azul, Lisbon, 2013, p.220

Photos 22 - 24: Recording with Aki Onda at ScratchBuilt Studios, Lisbon, 2014, p.223

PUBLICATIONS AND RELATED ACTIVITIES

Editorial and curatorial work:

Proceedings of INTER-FACE: International Conference on Live Interfaces (ICLI 2014), ed. Adriana Sa, Miguel Carvalhais, Alex McLean, pub. Porto University, CECL & CESEM (NOVA University), MITPL (University of Sussex), 2015. ISBN 978-989-746-060-9 http://www.liveinterfaces.org/proceedings2014.html

Adriana Sa, Joel Ryan, Edwin van der Heide, Atau Tanaka, Andrew McPherson, Thor Magnusson, Alex McLean, Miguel Carvalhais, Mick Gierson "Live Interfaces: Seeds of Debate". Preface discussion, in *Proceedings of INTER-FACE: International Conference on Live Interfaces (ICLI 2014)*, ed. Adriana Sa, Miguel Carvalhais, Alex McLean, pub. Porto University, CECL & CESEM (NOVA University), MITPL (University of Sussex), 2015. ISBN 978-989-746-060-9
http://www.liveinterfaces.org/proceedings2014.html

Journal Articles:

Adriana Sa, "Exploring Disparities between Acoustic and Digital Outputs", in *Leonardo, Vol. 48, No. 3,* MIT Press (1915): 280-281. doi:10.1162/LEON_a_01009
http://www.mitpressjournals.org/doi/abs/10.1162/LEON_a_01009?journalCode=leon#.Vp-kohiLSP8

Adriana Sa, Baptiste Caramieux and Atau Tanaka, "The Fungible Audio-Visual Mapping and its Experience", in *Journal Of Science And Technology Of The Arts 6, No. 1*, (2014): 85-96. ISSN: 1646-9798. doi: http://dx.doi.org/10.7559/citarj.v6i1.131
http://artes.ucp.pt/citarj/article/view/131

Adriana Sa, "How an Audio-Visual Instrument Can Foster the Sonic Experience", in *Live Visuals,* eds. L. Aceti, S. Gibson, S. M. Arisona, O. Sahin, *Leonardo Almanac 19, No. 3*, MIT Press (2013): 279-299. ISSN: 1071-4391
http://www.leoalmanac.org/vol19-no3-how-an-audio-visual-instrument-can-foster-the-sonic-experience/

Conference Papers:

Adriana Sa, Baptiste Caramieux and Atau Tanaka "A Study About Confounding Causation in Audio-Visual Mapping", in *Proceedings of XCoAX 2014* (Porto, 2014): 274-288. ISBN: 978-989-746-036-4 . http://2014.xcoax.org/pdf/xcoax2014-Sa.pdf

Adriana Sa, "Repurposing Video Game Software for Musical Expression: a perceptual approach", in *Proceedings of New Interfaces for Musical Expression* (London, 2014): 331-334. http://nime2014.org/proceedings/papers/343_paper.pdf

Adriana Sa, "Architecting the Audio-Visual to Modulate Rather Than Obfuscate The Sonic Experience", *Proceedings of the 1st International Conference on Live Interfaces* (ICLI 2012), Leeds (2012).
http://lipam.lurk.org/index.php?conference=LIPAM&schedConf=lipam2012&page=schedConf&op=presentations

Conference Organisation and Related Activities:

Steering Committee member of *ICLI - International Conference on Live Interfaces 2016,* University of Sussex.

Paper Committee member of the *41st ICMC-International Computer Music Conference*, University of the Arts Utrecht, 2016.

Paper Committee member of the *41st ICMC-International Computer Music Conference*, University of North Texas, 2015.

Program Committee member of *ICLC-International Conference on Live Coding* (papers and performances), University of Leeds, 2015.

Chair and producer of *INTER-FACE, the 2nd ICLI - International Conference on Live Interfaces*, Lisbon. Papers and discussions at Fine Arts University, installations at Chiado Museum, Performances at ZDB and National Music Conservatory, workshops at Institute of Art and Design. Lisbon, 2014.

Scientific Committee member and paper panel of *Conference Invisible Places*, Viseu, 2014.

Co-organization (with Adam Parkinson and Atau Tanaka) of the symposium *Real-time Visuals Workshop 2*, Goldsmiths College, London, 2013.

Co-chair of Master Students Seminar (Term 3), Creative Computing, Goldsmiths, University of London, June 5th 2014.

Reviewer for the *Conference on Human-Computer Interaction* (CHI 2013), Paris 2013.

Conference Talks:

Adriana Sa, "A Study About Confounding Causation in Audio-Visual Mapping", XCoAX 2014 , Porto, Portugal 2014. http://2014.xcoax.org

Adriana Sa, "A Perceptual Approach to Instrument Design and Composition", Tyneside Cinema, Newcastle, UK, 2014. http://www.realtimevisuals.org/conference/

Adriana Sa and John Klima, "Diverging Perspectives on Music with Video game Technologies". Workshop: Live Visuals for Performance, Gaming, Installation, and Electronic Environments, Northumbria University, February 26th, 2013 http://www.realtimevisuals.org/workshops/workshop1/

Adriana Sa, "Perception, Attention and Intensity in Performance", Ciclo de Conversas, Engineering Faculty, University of Coimbra, Portugal, April 10th, 2012.

Adriana Sa, "Architecting the Audio-Visual to Modulate Rather Than Obfuscate The Sonic Experience", International Conference on Live Interfaces (ICLI), Interdisciplinary Centre for Scientific Research in Music (ICSRiM), University of Leeds, UK, September 7th, 2012.

Performances:

The AG#2 software was premiered November 13th 2013. Performances were at:

Museu Municipal de Angra do Heroismo, Açores, Portugal, October 2016.

Sonic Writing symposium organised by Thor Magnusson, STEIM, Amsterdam, Netherlands, May 2016.

Carpe Diem, Lisbon, Portugal, July (with Nuno Torres and Nuno Mourão) and October (with Helena Espvall), 2015.

Invisible Places 2014, Viseu, Portugal, July 2014.

Teatro Nacional Maria Matos (with John Klima and To Trips), Lisbon, Portugal, March 2013.

EAVI performance series at Amersham Arms, London, UK, September 2013.

Goldsmiths College, London, UK, 2013.

The Arpeggio-Detuning software was premiered during the summer 2013. Performances were at:

Trem Azul, Lisbon, Portugal (with Joanne Cannon) 2013

Sonoscopia, Porto, Portugal (with Joanne Cannon) 2013

Sabaduo, Lisbon, Portugal

175 Santana, Lisbon, Portugal (with about 20 different musicians)

The AG#1 software was developed along with the Windowmatter project series (2008-12). Performances were at:

Calouste Gulbenkian Foundation (with John Klima), Lisbon, Portugal

Teatro S. Luis (with John Klima), Lisbon, Portugal

ZDB (with John Klima), Lisbon, Portugal

Canto do Século (with John Klima), Lisbon, Portugal

Festival Bang (with John Klima), Lisbon, Portugal

Musicbox (with John Klima) , Lisbon, Portugal

Espaço (with John Klima) , Lisbon, Portugal

Largo do Século (with John Klima, Toshio Kajiwara, O.blaat, David Maranha, Manuel Mota, Sei Miguel, Fala Mariam, Andre Gonçalves, Cesar Burago), Lisbon, Portugal

Regueirão dos Anjos (with John Klima, Toshio Kajiwara, O.blaat, David Maranha, Manuel Mota, Sei Miguel, Fala Mariam, Andre Gonçalves, Cesar Burago), Lisbon, Portugal

Moagem do Fundão, Fundão, Portugal

Bomba Suicida (with John Klima), Lisbon, Portugal

St. James Church, Goldsmiths University, London, UK

Off-Limits, Madrid, Spain

Share, NYC, US

DOCUMENTATION AND SUPPORT MATERIAL

URL about this research: http://adrianasa.planetaclix.pt/research/research.htm

Mappings used in the Study on Audio-Visual Mapping and Perception (4’27”):
http://doc.gold.ac.uk/~map01apv/StudyMappingPrototypes.mp4

Site about the AG#1 software, including screen capture with zither input (6’11”):
http://adrianasa.planetaclix.pt/research/AG1.htm

Site about the Arpeggio-Detuning software, including solo and collaborative recordings:
http://adrianasa.planetaclix.pt/research/ArpeggioDetuning.htm

Solo recordings with zither & Arpeggio-Detuning software:
Zither + audio sample bank 1 (2’19”):
http://eavi.goldsmithsdigital.com/wp-content/uploads/2013/10/adri_set_1.mp3
Zither + audio sample bank 2 (2’30):
http://eavi.goldsmithsdigital.com/wp-content/uploads/2013/10/adri_set_2.mp3
Zither + audio sample bank 3 (4’06”):
http://eavi.goldsmithsdigital.com/wp-content/uploads/2013/10/adri_set_3.mp3

Collaborative recordings with zither & Arpeggio-Detuning software:
With Joanne Cannon (37’50”): http://www.cityarts.com/audio/sa&cannon.mp3
With Helena Espvall (18’45”): http://www.cityarts.com/audio/sa&espvall.mp3
With Tó Trips (17’03”): http://www.cityarts.com/audio/sa&trips.mp3
With Aki Onda (22’55”): http://www.cityarts.com/audio/sa&onda.mp3

Site about the AG#2 software, including screen captures with zither input (6’46” and 4’36”) and performance (4’47”): http://adrianasa.planetaclix.pt/research/AG2.htm

CHAPTER 1

Introduction

1.1 An Artistic Position

This research endeavours to clarify insights derived from artistic practice. Prior to introducing the research questions, this first section of the thesis introduces a personal creative position.

1.1.1 A Personal Context

This research is motivated by creative concerns that can be traced back to the beginnings of my artistic practice. I always thought that music involves much more than sound. During my six years of classical piano training as a child, I was advised to imagine different characters and situations for each part of the music, and develop the expression of my playing accordingly. During my visual arts education, my mentor, an artist and Zen practitioner, taught us that "making for the sake of making" is a fundamental artistic principle, regardless of the medium. On the one hand, we needed to discover the reality beyond our usual ways of perceiving the world. On the other, we needed to permeate that wider reality with our works, so that the audience could have their own, individually significant experiences. Armed with this understanding, my thinking about painting started to fluidly embrace volume, space, sound and performance. At some point I realised that my thinking was actually about music.

In the late 1990s I started to use sensor technologies and software for relating sound with light, space, movement, architecture, weather and social context. During an installation in the Japanese Northern countryside, a musician stretched an amplified wire from my gallery space to the forest, so as to integrate the sound of the snow falling with the soundscape of my installation, which changed according to daylight. I started playing the amplified wire; it felt very similar to piano playing, but freed from any concrete musical tradition. I multiplied the wires and developed the "bodiless instrument", which acquired different timbrical qualities in every performance-installation. As I moved in space to play the wires, my shadow upon light sensors also affected the processing of digital sounds. At the time I used MIDI, which means that every input signal was scaled to a maximum of 127 digital values, regardless of the sensor sensitivity and/ or the resilience of the input. It did not feel like a limitation, inclusively because each of those values could be mapped to any number of parameters. Neither the digital nor the acoustic components of my setup were fully controllable. I found this stimulating in performance, because the interaction with the system felt reciprocal, and I needed a particular state of concentration to create musical meaning upon the unexpected.

Later I felt that I wanted more control over the musical output, yet simultaneously, I did not want to dispense with unpredictability. The acoustic sound became gradually more controlled when I started customising existing multi-string instruments. In addition, for a while I projected video images upon light sensors so as to process digital sounds. However, I found that the musical results were not particularly rewarding, and also, I was not really interested in video. I needed a new system, but how could it connect the acoustic and the digital sound? And how could it connect the sound with a digital moving image? We are used to VJ culture, where rapidly changing pictures are constantly synchronised with dance beats. I wanted to invite the audience for a very different kind of experience. But how could they feel invited to focus upon the subtleties of music, and how could I extend the psychological space of the work beyond the physical performance space?

1.1.2 A Personal Audio-Visual Instrument

The audio-visual instrument developed in this research combines acoustic sound, digital sound, and digital image. It includes an acoustic instrument called a *concert zither*: a German-Austrian multi-string instrument with a resonant body and a fretboard. This zither has custom strings and tuning. Its sound is used as an input to the audio-visual software. The software processes 3D graphics and recorded sounds. Sound and image affect each other reciprocally, in ways that will be described later in the thesis. In performance I use two interfaces: the audio analysis of the zither, and a set of switches (from a game pad, an iPad, or a computer keyboard).

The instrument utilises software that is originally intended for the creation of video games. Video game engines can render a digital 3D world, and make the position of elements in that world have an impact upon the spatialisation of related sounds. But as I started to work with video game engines, I needed to question certain theories embedded in the digital platform. For example, the software assumed that digital behaviours must be consistent, and cause-effect relationships clearly perceivable. Those theories were concealed in the code, and I was able to implement my modifications with the collaboration of John Klima, expert in video game programming, C and C++. This entailed many discussions about technical feasibility, and I realised that I had to clarify the boundaries of my creative motivations. I needed clear design principles to specify digital functions and constraints.

Software mediates physical action through code, and code embeds theories informed by specific purposes and criteria. The problem is, those theories are too often taken for granted. In everyday life we are used to handling computers as magic black boxes that save us labour. When the black box works, its origins are forgotten. The more science and technology succeed, the more opaque and obscure they become, and the more distant we become from computation as creative material.

Murray-Browne states that "there is a conflict underlying many novel musical instruments in that they are intended to be both a tool for creative expression and a creative work of art in themselves, resulting in incompatible requirements" **[**Murray-Browne et al. 2011**]**. This might not be the case. An instrument indicates a hierarchy of concerns, which determine its functions and meanings. To the audience some can be explicit and others implicit, or even hidden. With novel instruments, inexplicitness and equivocalness can be as important as in any artwork. For example, the actual input of certain musical instruments is concealed to the audience - e.g. body-synapses, ultrasounds or infrareds, as explored by Sensorband[1]. Yet the performers' body gestures and the instrument setup can indicate a relation with unperceivable phenomena. This makes the audience seek and imagine how the system works. The absence of explicit information leads to interrogation, which in turn influences perception.

My audio-visual instrument has a level of significance that is suggested but not explicit. It suggests that my creative concerns are incompatible with certain theories governing video game software. The instrument can be linked with video gaming because there is a 3D world and the virtual camera dynamics are synchronised with input detection, as if the zither were a game pad. At the same time, the qualities of sound and image, the seemingly inconsistent audio-visual mapping and the performance setup make the instrument stand apart from gaming.

1.2 Research Aims and Questions

The thesis formulates a perceptual approach to audio-visual instrument design, composition and performance. It draws from several fields of research, such as psychology, neuroscience, interaction design and music. The approach leads to a set of creative principles and a related parametric visualisation model. The model can be used to analyse

[1] http://www.ataut.net/site/Sensorband and http://www.youtube.com/watch?v=o-ZcsAHVn6A

sensory dominance, sonic expression and spatial presence in any audio-visual performance language. Intertwining theory and practice, the research leads to the expansion in depth of a personal audio-visual performance language.

The primary concern of my creative practice is to develop audio-visual performance work where the audience can focus upon the subtleties of music. The interaction with my instrument should convey complex sonic constructions. The image should extend the physical performance space to a digital 3D world beyond the screen. It should function as a stage scene, which reacts to the sound without distracting attention from the relation between the sounds themselves.

As Michel Chion observed, the combination of sound and image generates a third, audio-visual element [1994]. An important question here is if, and how it can benefit the experience of music. From Pierre Schaeffer [1966] to recent acousmatic composers, many people have argued that sounds must be detached from their originating cause to be fully experienced. Jeff Pressing also investigated the audio-visual relationship in digital 3D environments, noting that perception operates from vision to audition whenever a direction of causation is discernible [1997:8]. Furthermore, Meghan Stevens assessed that the music remains dominant in audio-visual performance when the audio-visual relationship is partially congruent, yet she is the first to admit that her theories rely on limited evidence [2009:3]. Indeed, visual dominance is a challenging problem. But in neuroscience, Sinnett et al. found that attention can be manipulated so that vision does not dominate over audition [2007]. My initial creative purpose is to not let vision dominate over audition, which requires manipulating the audience's experience. Subsequently, my intention is to make this less obvious, and explore gray areas so that the audio-visual relationship produces a greater variety of perceptual effects. In other audio-visual performance languages, the purpose of the image can be very different from the one in my work. Clarifying the problem of sensory dominance should also help to clarify those different purposes.

The question of sensory dominance might be the most challenging problem in this research, yet other problems are equally important. Assessing how interaction can convey sonic expression is crucial for the design and configuration of any music performance system. In my practical work, the performer's control over the instrument and the instrument's unpredictability should be calibrated so as to convey "the awaken of unused abilities", using an expression employed by Schroeder and Rebelo [2009]. Such purpose is not uncommon in instrument design. Pinning down a corresponding notion of sonic expression should also help to clarify how other notions of expression inform different types of interaction design.

To estimate how the combination of sound, image and physical setup inform the audience's sense of spatial presence is useful as well. Spatial presence can be defined as "the subjective experience of a user or onlooker to be physically located in a mediated space" [Hartmann et al. 2015:117]. The psychophysical space of an audio-visual performance is a mediated space. The work can explore perceptual cues and manipulate attention in order to drive the audience's sense of presence. My practical work oscillates between drawing attention to the performer and away from the performer, to the environment and to imaginary spaces beyond the physical performance space. It requires me to consider spatial presence with respect to sensory dominance and sonic expression. Beyond personal practice, assessing how stimuli inform spatial presence is beneficial to any audio-visual performance work.

The thesis considers audio-visual performance in an integrated manner, relating aspects relevant to the performer's real-time creation process and the audience's experience. The research addresses several interrelated questions.

The first set of questions aims at defining personal creative principles for the sound, the image and the audio-visual relationship, independently from technical means. How can an instrument design convey sonic expression? How can the relationships between the sounds be fully experienced, while the audio-visual relationship generates additional effects and meanings?

The second set of questions aims at the creation of a parametric visualisation model. The model should reveal how any audio-visual performance work may converge and diverge with my creative principles. In addition, it should provide cues about physical setup and spatial presence. So how can we parameterise sonic expression? How can we parameterise sensory dominance? How can we parameterise the physical setup and the semantics of a work? And how can the model indicate the audience's sense of spatial presence?

The third set of questions aims to explore my principles for the sound, the image and the audio-visual relationship in instrument design, composition and performance. How can the combination of direct interaction with an acoustic instrument and indirect interaction with software convey complex musical constructions, which exceed the performer's deliberate intention? How can a digital 3D world function as a stage scene that reacts upon the sound, and how can it allow for the audience to focus on the subtle nuances of the sonic construction? How can the combination of sonic/visual materials, audio-visual mappings, physical setup and sound diffusion intertwine the physical and the digital 3D space? And how versatile should the work be?

1.3 Starting Points and Contributions

This section outlines the contributions of this research, situating where the questions leading to those contributions come from. I will mention creative motivations, related theoretical work and adopted research methods, but without detailing any aspect. The details are provided in chapter 2, which presents a literature review and discusses research methods.

The previous section presented three chaining sets of research questions, which lead to three types of contributions. The first type consists of three creative principles, for the sound, the image and the audio-visual relationship. These are presented in chapter 3. The second type is a parametric visualisation model, which can be used to analyse any audio-visual performance with respect to sonic expression, sensory dominance and spatial presence. This is presented in chapter 4. The creative principles and the parametric model inform the practical contributions of this research, presented in chapter 5.

1.3.1 The Advantage of a Parametric Visualisation Model

Parametric models are informed by particular research frameworks, which open particular windows of discussion. The one of this thesis relates to previous models, which focus on human-computer interaction [Birnbaum et al. 2005, Magnusson 2010] and electronic music performance as audio-visual experience [Ceciliani 2014]. This new model is unique in parameterising sonic expression, sensory dominance and spatial presence. It reflects concerns that govern my personal practice - the development of an audio-visual instrument and performance language, as well as decisions about the physical performance setup. Yet, each artistic language has a reason-to-be, and in scientific research it is not very useful to promote one sphere of concerns over another. The model does not aim to promote a personal artistic sensibility, as much as it aims to provide an operational means for analysing and developing any audio-visual system. I will discuss performances by Steina Vasulka, Ryoji Ikeda, and Thor Magnusson, demonstrating that the model can be used to analyse a diversity of audio-visual performance languages, be they concerned with sensory dominance, or not at all.

1.3.2 Sonic Expression: Creative Motivations & Theoretical Contributions

Sound organisation plays a central role in my practical work, which explores musical forms with subtle chromaticisms[2] and timings. The aesthetic approach of the sonic mappings combines acoustic and digital sounds based on tonality, timbre, semantics and morphology. The zither sound intertwines with sounds of nature, musical instruments, and synthesised sounds. My sonic constructions cannot be labelled according to any particular style, but shifting in-between tones is an ancient musical practice, inherent to a variety of musical idioms and traditions such as the Portuguese *fado* or the Indian *raga.* In the words of a Hindustani classical musician, *raga* vocalists often “move between pitches freely and render the inter-tonal spaces carefully (…) Performers often use concepts such as the approach towards the note, the release of it, and so forth” [Kelkar and Indurkhya 2014]. Such expressive details emerge with intuitive precision and cannot be scored. They bring life to the music. Allowing for intuitive precision is important in my instrument design, which seeks to convey a myriad of expressive shifts, be they chromatic, timbrical or morphological.

The interaction with my instrument entails the performer controlling the acoustic output, but not fully predicting the digital output. This approach to interaction design relates to literature about the epistemic dimension of digital tools [Magnusson 2009], the compositional role of “chance” [Cage 1961], the role of effort in music [Ryan 1991], and the performer’s state of perception [Schroeder and Rebelo 2009].

The thesis bridges these ideas with perception science [e.g. Snyder 2001, Knudsen 2007], and examines how different types of interaction design convey different understandings of sonic expression (chapter 2.5). I develop one particular understanding, elaborating on how the two-folded, reciprocal interaction between performer and instrument conveys expression (chapter 3.3). The investigation leads to two contributions related with sonic expression. One is a creative principle: to threshold the performer’s control over the instrument and the instrument’s unpredictability (chapter 3.7). The other is a way to

[2] Chromaticism is a compositional technique interspersing the primary diatonic pitches and chords with other pitches of the chromatic scale.

parameterise how different notions of sonic expression and different interface designs might converge or diverge from this principle (chapter 4.3.2 and 4.3.5).

1.3.3 Sensory Dominance: Creative Motivations & Theoretical Contributions

Another set of contributions, perhaps the most important, respects to the artistic debate on sensory dominance. The question is if, and how the relationships between the sounds can be fully experienced while the audio-visual relationship generates additional effects and meanings. Chion observed that the mutual impact of vision and audition leads to a perceptual surplus [1994]. However, the strength of each sensory modality in the perceptual representation tends to be unbalanced. The problem of visual dominance has been debated in music [e.g. Schaeffer 1966, Lopez 2004] and audio-visual theory [Pressing 1997, Stevens 2009]. It lacks resolution.

To clarify the problem, the thesis bridges audio-visual theory and scientific research on perception, multi-sensory integration and attention (chapter 2.3 and 2.4). It looks at how perception is driven to make sense of multi-sensory complexity, considering conscious and unconscious information processing. It investigates why some stimuli attract automatic attention and others do not, and why both automatic and deliberate attention influence how stimuli are perceived. And, it looks at what perceptual binding implies in terms of sensory prioritisation.

From these investigations, I extrapolate a method to analyse the sonic and the visual dynamics, and compare their perceptual strength (chapter 3.4). Additionally, I extrapolate a method to analyse how different ratios of congruence/ incongruence influence perceptual prioritisation (chapter 3.5). This is complemented with a study on audio-visual mapping and perceived causation (chapter 3.6), whose conclusions are equally applicable to the audio-visual relationship in space.

The research about sensory dominance culminates with several contributions: two creative principles (chapter 3.7) and a method to parameterise and analyse sensory dominance in any audio-visual performance (chapter 4.3.3 and 4.3.5). One creative principle dictates that visual dynamics should dispense with abrupt discontinuities, so that that the image never becomes protagonist. My practical work seeks to explore the principle in ways that make the image appear organic. The other creative principle dictates that the audio-visual relationship should produce a sense of causation, and simultaneously confound the cause and effect relationships. My practical work seeks to explore the principle in audio-visual mapping and physical space, considering sound spatialisation and physical gesture.

1.3.4 Spatial Presence: Creative Motivations & Theoretical Contribution

In my work, the moving image is projected over the performer, extending the psychological space of the work to a digital 3D world that morphs with sound. At the same time, the sonic construction oscillates between creating environmental soundscapes, or highlighting the performer's playing, or evoking imaginary natural spaces. The work seeks the audience to feel immersed in the music, and their sense of spatial presence should embrace the physical and the digital space.

The sense of spatial presence has been discussed in sound art [Lopez 2004, and Voegelin 2010] and media theory [e.g. Sacau et al. 2008, Hartmann et al. 2015]; from these works we can infer that spatial presence is influenced by individual predispositions, the characteristics of sound and image, and the audio-visual relationship. Other researches discuss how the psychological space of the work is informed by speaker placement [e.g. Chion 1997, Emmerson 2007, Ciciliani 2014]. Also, the psychological space of a work is inextricably related with the semantics of that work. Pressing proposed a useful semantic characterisation of sounds [1997] that can be extended to the audio-visual domain.

Regarding the topic of spatial presence the thesis' contribution consists of drawing a relation between the abovementioned works (chapter 4.3.4 and 4.3.5). The parametric visualisation model proposes that spatial presence can be exploited from the combination of interaction, sonic and visual dynamics, audio-visual relationship, physical setup and semantics. In addition, a high-level parameter representing the psychological space of the work provides cues about factors that cannot be directly represented in the model.

1.3.5 Practical Work

The perceptual approach of the thesis guides the discovery and development of an audio-visual performance language, described in chapter 5. The parametric model's level of abstraction is appropriate to inspire the expansion of creative strategies. Yet those strategies would not have been discovered and developed without long periods of studio practice, and many performances.

The previous subsections mentioned three creative principles for the sound, the image, and the audio-visual relationship. They can be explored in many different ways. My instrument explores the principles with a custom zither and audio-visual software, which operates based on the zither sound. My creative strategies take advantage of disparities between the direct acoustic output of the zither, and the indirect outputs of digital sound and image. The first version of the instrument emphasises a few ‘chaotic’, non-linear features in the zither, exploring digital configurations and mappings that cause the software to behave inconsistently. The final version implements compositional strategies at low level in the digital architecture, rather than solely at high level in parameterisation. The difference value between the detected pitch from the zither and the closest tone/ half tone is applied to the processing of digital sounds and to the audio-visual mappings. Regarding the visual dynamics and the question of sensory dominance, my creative decisions reflect that I am more driven to face gray areas as creative material than to manipulating the audience’s experience. Furthermore, both versions of the software use functions characteristic of 3D game engines, so that visual dynamics affect sound spatialisation. However, the strategy for sound diffusion is uncommon with 3D environments, as an inverted stereo system blurs the correspondence between the image and the physical sound source.

Considering my creative principles and parametric visualisation model, the work evolves with respect to interaction, dynamics and semantics of sound and image, audio-visual mapping, digital sound spatialisation and sound diffusion.

1.4 Thesis Structure

Chapter 2 provides a literature review, complemented with a discussion of research methods. The first section of the chapter provides a general context for my practical work, reviewing the notion of instrument as an open-ended system, the implications of physical vs. digital mediums, and the notion of spatial presence. This provides perspective over the following sections. The chapter proceeds with a review of audio-visual relationships and artistic debates about sensory dominance. These debates motivate a review of perception science literature, which is equally useful to distinguish how different understandings of sonic expression might inform different types of interaction design.

Chapter 3 draws from the previous reviews, extrapolating theoretical tools that are useful for the design of the parametric visualisation model and the practical work of the thesis. The chapter presents a way to analyse the relation between interaction and sonic expression. It presents a way to analyse the sonic and visual dynamics, and compare their relative strength. And, it presents a way to analyse how different ratios of audio-visual congruence/ incongruence affect perceptual prioritisation. This is complemented with a study on audio-visual mapping and perception, whose conclusions also apply to the audio-visual relationship in space. The investigation leads to three creative principles, which respect to the sound, the image, and the audio-visual relationship.

Chapter 4 presents a parametric visualisation model for audio-visual instrument design, composition and performance. Prior to describing this model, the chapter analyses previous models, identifying useful aspects and problems. The model of the thesis is used to analyse how different audio-visual performance languages converge and/or diverge from the creative principles that govern my practical work. On the one hand, the investigation exposes alternative ways of exploring each principle. On the other, it shows that the model is useful to the analysis of any audio-visual performance practice.

Chapter 5 applies the previous investigations to creative practice. It describes the first version of my audio-visual instrument and the physical performance setup, considering my creative principles and the study on audio-visual mapping and perception. Subsequently, the chapter describes an expansion of creative strategies in depth. The expansion begins with a focus on sound organisation, involving the creation of new audio software. Compositional strategies are then adapted and extended with 3D audio-visual software. The investigations unfold through visual dynamics, audio-visual mapping, digital sound spatialisation and sound diffusion quality.

Chapter 6 discusses the contributions of this research, including the relation between my practical work and the parametric visualisation model. It begins with a discussion of how the model parameterises sonic expression, sensory dominance and spatial presence, considering my creative principles and how this research can contribute to the scientific creative community. Then I use the model to discuss my solo performance work, outlining a possible difference between the performer's and the audience's experience. Subsequently, I define a method to analyse and represent the work's versatility in collaboration with other musicians. The chapter presents four case studies, demonstrating that the model can reveal how the instrument seeks to enable a diversity of audio-visual performances, and simultaneously put limits to this diversity. The closing section of the chapter consolidates the discussion of the thesis, with a summary of contributions.

CHAPTER 2

Literature Review and Research Methods

2.1 Chapter Introduction

This chapter reviews practical and theoretical work relevant to the questions and concepts that lie at the centre of this research. The chapter unfolds as follows:

Section 2.2 provides a general context for my practical work. The section introduces the notion of instrument as open-ended system. It distinguishes physical and digital mediation, pointing out relevant perspectives upon interaction and perceived causation. It looks at how spatial presence has addressed in film, media and electronic music performance, providing perspective over how it can be considered in audio-visual performance.

Section 2.3 focuses on audio-visual relationships. It reviews a set of audio-visual performance and animation works, distinguishing two types of approaches. It proceeds with an overview of works exploring music with digital 3D environments. It finalises with a review on how sensory dominance has been debated in music, film and audio-visual theory.

Section 2.4 looks at psychology and neuroscience so as to clarify the problem of sensory dominance for music perception; it establishes the scientific foundations of the perceptual approach developed in the thesis. It reviews research on perception, multisensory integration and attention, as well as definitions of 'intensity' in music.

Section 2.5 focuses on interaction and sound organisation, bridging artistic views and perception science. The section reviews how data can be mapped to meaning, and the notion of embodiment. It situates a personal notion of sonic expression with literature in music, interaction design and psychology; and it looks at how different understandings of flow substantiate in interaction design.

Section 2.6 discusses the methods of the thesis.

2.2 Instrument and Spatial Presence

This section situates my practical work in a way that provides perspective over the following sections, which focus on audio-visual relationships and sonic expression in detail. The current section reviews the notion of instrument as open-ended system. It outlines a difference between physical and digital mediation, and looks at a set of works emphasising the epistemic dimension of digital tools; namely at how the authors face interaction and perceived causation. Furthermore, the section reviews how the sense of spatial presence has been addressed in film, media and music.

2.2.1 Instrument as Open-Ended System

We can draw a separation between instrument, composition and performance, but in practice that separation might not be so obvious. Thor Magnusson describes 'composing an instrument' as defining and limiting the boundaries of a musical space to be traversed in performance [Magnusson 2010]. The term is extended by Tim Murray-Browne [Murray-Browne et al. 2011], who proposes an approach to instrument creation as an art form in itself, where instrument, mapping and music are an integrated part of a greater composition. Indeed the instrument developed in this research is not solely a tool, but creative work itself. Its design and physical setup are informed by compositional strategies that consider the performer's and the audience's experience.

The term 'instrument' can refer to a "self-contained and autonomous sound-producing object that enables a musician to perform in a live situation" [Tanaka 2009:236] - an acoustic instrument or a non-acoustic instrument such as the electric guitar and the synthesizer. The term can also refer to an open-ended system [Tanaka 2009]. The instrument is then not a single object because it includes several components. It is also not really self-contained because it depends on content, context and configuration. For example, with Pierre Schaeffer and *musique concrète* the studio became the instrument of the composer. And *turntablism* appropriates the turntable as an instrument, relying on performance techniques and pre-recorded materials. For artists like Steina Vasulka, moving images are also part of the instrument; her instrument includes a digitally adapted violin and video software[3]. Furthermore, the quality and size of the amplifiers and speakers also have an enormous effect on the final sound, and artists like Aki Onda consider them a part of the instrument[4] as well.

Atau Tanaka described how the metaphor of "musical instrument" applies to a musical and instrument-building tradition that is often labelled NIME, after the international conference series New Interfaces for Musical Expression, which started in 2001 [2009]. Tanaka describes these instruments as open-ended systems, which depend on content, context and configuration. Typically, they comprise: an input device (e.g. sensor and data

[3] See for example "MIDI violin demo" (1998) at http://www.fondation-langlois.org/html/e/page.php?NumPage=419
[4] http://www.akionda.net/

acquisition subsystem); mapping algorithms (software subsystem translating incoming data into musical information); parameters modulated by live input; a structural layer defining musical sections; and an audio output subsystem consisting of output channel configuration and digital-to-analogue converters (DAC). Tanaka also points out that an acoustical instrument can be thought of as data to be "sensed" and analysed. The application of mapping techniques turns the microphone into an interactive sensor.

The extent to which NIME designs allow for agency is quite variable. For example, in an article titled "Beyond Guitar Hero - Towards a New Musical Ecology", Tod Machover declares the purpose of "diminishing the current exaggerated distinctions between celebrities and amateurs", so as to compensate for "people's limitations" [2008]. Such "compensation" implies that the software prescribes which output results are desirable, and which are not. In opposition, one can argue that an instrument should not compensate for limitations: an idiosyncratic, personal interface, which requires particular skills and constant practice, can be governed by very different principles than interfaces intended for non-musicians. Pinning down how interface designs condition musical expression is an important matter of research in the thesis.

2.2.2 Instrument as Mediation

Mediation is a central aspect of all creative interfaces, and an instrument combining acoustic and digital components affords two fundamentally different types of interaction. Thor Magnusson points out that when the performer yields physical force to drive an acoustic instrument, both the instrument body and the sound engine behave according to physical properties [2009]. Conversely, in digital instruments the physical force "can be mapped from force-sensitive input devices to parameters in the sound engine, but that mapping is always arbitrary (and on a continuous scale of complexity)". Magnusson calls this the epistemic nature of software.

The epistemic nature of software can be explored in many different ways. I chaired the second International Conference on Live Interfaces (ICLI 2014), and took the opportunity to extend the proceedings with a preface interview/ discussion including contributions by Joel Ryan, Edwin Van der Heide, Atau Tanaka, Mick Grierson, Andrew McPherson, Miguel Carvalhais, Magnusson and Alex McLean [Sa et al. 2015]. The interview explores several aspects important to this research, and I will refer to it several times in this chapter. In this early section I will quote contributions that highlight how developing an instrument with software implies assuming a position regarding the performer's control over the instrument, and the audience's understanding of the functional cause and effect relationships.

McLean and Magnusson are part of the live coding movement, which has a particular way of emphasising the epistemic dimension of software. Live coding is a practice where software that generates music and/or visuals is written and manipulated as part of the performance [Collins et al. 2003]. Usually, the code is projected on a screen so that people can see the process. According to McLean,

> (…) I think there are political reasons for projecting. Not too long ago the fashionable movement for creative coding was 'generative art' (...) Generative artists have endless discussions about authorship -- if you program a computer to make art, is the author the programmer, or the computer? In my view this whole question of authorship is an intellectual cul-de-sac; humans have always thought through their tools, and followed lines through their materials. Thankfully live coding makes this question redundant, no-one can deny the human influence in such a performance. [Sa et al. 2015]

The practical work developed in this research is not live coding, but it entails a related political aspect. It seeks to emphasise the human behind the machine, while stressing that one's interaction with a digital system is always mediated by theories embedded in multiple layers of code, regardless of whether the interaction feels immediate or not.

When Magnusson was asked about the amount of music theory encapsulated in his live code he wrote:

> The more low-level the programming language is, the more control you have over the hardware; the higher you get in this stratification, the more constrained you are by the abstractions defined by the system. But you gain speed: for a musician or an artist working with computers, the key question is at what level they want their constraints to be. (…) Time is always an important constraint as well (…) Personally I am interested in coding at a high musical level. [Sa et al. 2015]

Encapsulated music theory and speed are important for the instrument developed in this research, but the software is not intended to provide a sense of immediacy. Code can produce unexpected results, and immediacy comes rather from the direct, physical interaction with the zither. My instrument explores disparities between physical and digital mediation, and the thesis seeks to ground my creative choices while looking at the construction of musical time.

In the ICLI 2014 interview I also asked Magnusson, McLean and Carvalhais if they find it important that the audience understands their systems. Carvalhais (who is a not live coder but performs with a laptop) finds it important to provide functional clues, safeguarding:

> This doesn't mean that all the details of the process are understood, but just that the audience is able to predict relationships and thus be surprised whenever either system or interactor deviate from the predicted outcomes. [Sa et al. 2015]

Magnusson says:

> I don't think musical instruments should be necessarily easy to play or understand. We're not designing buttons in an elevator or a coffee machine where the affordances responding to the thing's function should be understood immediately. [Sa et al. 2015]

And McLean says:

> Understanding code is not important to me, in fact in *Slub* (a club band with Adrian Ward) we have sometimes purposefully obscured our code to make it more difficult to read, while still showing some of the activity of the edits. When I watch live coding performances, I don't read the code. Indeed even as a live coder I don't have top-down understanding of what my code is doing, I am just working with the code as a material, while listening to the output of the process it describes. [Sa et al. 2015]

Providing functional clues while obscuring how the system actually works is an important matter of research in this thesis. The thesis seeks to clarify how that affects the perceptual experience of music.

2.2.3 Spatial Presence and Performative Arena

Film and media can forge audio-visual relationships so as to create a sense of space. Michel Chion coined the term *superfield* to designate the sound space created by multi-track sound and multi-speaker placement in the movie theatre [Chion 1994]. This notion denotes the highly varied soundscape of “ambient natural sounds, city noises, music, all sorts of rustling”. The superfield does not depend on what we see moment by moment on screen; instead it provides “a continuous and constant consciousness of all the space surrounding the dramatic action” [1994:150].

In media theory, Draper et al. proposed that spatial presence depends on attentional processes [1998]. Schubert described spatial presence as *cognitive feeling*: “a feedback of unconscious processes of spatial perception that try to locate the human body in relation to its environment” [Schubert 2009:15]. Sacau et al. described spatial presence as a mental phenomenon informed by the properties of media stimuli, as well as by psychological factors such as the capability to be immersed [2008]. These authors concluded that in radio and TV, the sense of spatial presence depends more on the active suspension of disbelief than on the properties of media stimuli. Conversely, in complex interactive virtual environments designed to convey participatory interaction the sense of spatial presence would depend more on the properties of stimuli than on individual psychological factors, such as the capability to feel immersed. These conclusions might be arguable, but their starting point is solid: spatial presence depends on stimuli as well as individual predisposition. Stimuli can condition predisposition and attention to a lesser or greater degree. Hartmann stresses that spatial presence is compatible with disbelief: a person can feel simultaneously aware of their mental transportation and their physical location [2015].

3D video games are frequently mentioned as an example of how stimuli can convey spatial presence in a virtual environment. Paul Dourish stressed that interfaces apply the study of perceptual mechanisms so that the player feels present in the digital 3D world beyond the screen [2004]. An important functionality of digital 3D engines is 3D positional audio effects. 3D positional audio effects emerged in the 1990s, for PC and game consoles. They make sound spatialisation depend on visual dynamics. The creators of *FMOD Studio*, a popular 3D positional audio software, explain how this establishes the difference between 2D and 3D sound[5]. TV and CD recordings use 2D sound: each aspect of the sound environment is directed to a particular output speaker, regardless of whether there is one or many speakers (surround sound is also considered 2D). Conversely, with 3D sound the sonic output depends upon the position of the source in the digital 3D world, relative to the 3D camera/ listener. If the camera, and/or the sound emitting object in the digital world moves, the overall audio output reflects this shift. This is achieved by constantly altering how the signal is routed to the output channels. The audio-visual relationship mimics how we perceive the physical world, and that conveys spatial presence in the digital world.

Mimicking how we perceive the world seems also important when the desire is to emphasise a connection between the digital 3D environment and the physical space, as is the intent of a system developed by Mike Wozniewski, Zack Settel, and Jeremy Cooperstock [2006a, 2006b, 2006c]. Designed for participatory interaction, the system is controlled by regular physical activity, and the sending of sound signals from one area to another is based on physical models of sound propagation. Although the authors bend the rules of physics in order to explore the musical potentials of the work[6], the users' interaction relies on a clearly perceivable, congruent relation between gesture, sound and image.

Clearly perceivable cause and effect relationships are often not the goal in electronic music performance, as shown by Simon Emerson's distinction between *local functions* and *field functions* [2007:92]. Whilst local functions seek to extend, but not break, the perceived relation of human performer to sounding result, field functions create a context, a landscape or an environment where local activity may be found. Emmerson prefers not to

[5] see "3D Audio", in *FMOD Studio 101* (2014) :296
http://www.soundlibrarian.com/uploads/3/1/0/6/3106267/fmod_studio_101_chapter_10_3d_audio_gdc.pdf
[6] http://wikibin.org/articles/audioscape.html

play down this division, but add supplementary dimensions to it - local can become field, and vice-versa. He uses the term *performative arena* to describe the relation between performer, audience, space, sound sources and events. We can say that the performative arena relates the physical and psychological space of the work, driving the audience's attention and spatial presence. In audio-visual performance, this mediated space depends on the spatial relation between performer and visual projection, the speakers' placement, the characteristics of sound and image, and the audio-visual relationship. Accordingly, my practical work must consider sonic expression, sensory dominance and spatial presence in an integrated manner, from instrument design and configuration to performance setup.

2.3 Audio-Visual Relationships and Sensory Dominance

Music is the central aspect in my practical work, which considers the global perceptual experience, involving vision as well. My music making never felt a lack of scientific research, but that changed once I started to work with projected moving images. This section situates those initial creative concerns. It distinguishes two types of approaches to audio-visual performance and animation – those where all the sonic and visual events are synchronised, and those where they are not. Subsequently, it reviews works exploring music with digital 3D environments. And finally, it reviews how the problem of sensory dominance has been debated in music, film and audio-visual theory, including arguments concerned with synchrony, perceived causation and congruence.

2.3.1 Approaches to Audio-Visual Performance and Animation

There are many different approaches to audio-visual performance and animation, particularly if we generalise the terms so as to embrace all works that combine sound and image, with the exception of narrative film. If we abstract from this overwhelming diversity and look solely at how the timing of visual events relates with the timing of auditory events, we can distinguish two types of approaches. One type explores audio-visual relationships where *all* the sonic and visual events are synchronised one-to-one. I will call it the "one-to-one approach". The other type explores relationships where at least a part of the sounds and images are not synchronised. I will call it the "structural approach". This term should not be confounded with that which is known as structural film - the experimental film movement prominent in the US (1960s) in the UK (1970s), which is based on the structuralism theories from linguistics.

There are many examples of the one-to-one approach. Father Castel's *Ocular Harpsichord* (1730) consisted of a harpsichord with coloured glasses and curtains; when a key was struck, a corresponding curtain would lift briefly to show a flash of corresponding colour [Moritz 1997]. Synchrony was prescribed by these technical characteristics.

From a technical perspective, Wallace Rimington's colour organ the *Clavier à Lumières* (1890s) allowed for synchronised and non-synchronised audio-visual relationships: it had two keyboards, one to play the music and one to "play colours"; and it was possible to set up the device in such a way that the rainbow scale was synchronised to the notes time [Rimington 1895]. In any case, Remington was inclined to one-to-one synchronisation: "both sound and light may be conveniently played at the same" [1912:15]. He established direct correspondences between sounds and colours, and endeavoured the colour-keyboard to reflect the note scale.

Another example of the one-to-one approach is the *Variophone* created by Evgeny Sholpo (1930)[7]. It was an optical synthesiser, which utilised sound waves cut onto cardboard disks rotating synchronously with a 35mm film, while being photographed onto it to produce a continuous soundtrack. This filmstrip was then played as a normal movie. Read by a photocell and amplified by a loudspeaker, it functioned as a musical recording process. Again, synchrony was prescribed by technical characteristics.

The abstract film work of Norman McLaren *Dots* (1940)[8] is also a clear example of the one-to-one approach. McLaren was one of the most influential animators of the last century. He painted directly onto clear frames of film. The music was created in the same way, painting directly into the area on the filmstrip usually reserved for the soundtrack.

Dafne Oram's *Oramics Machine* (1957)[9] seems to convey both the one-to-one and the structural approach. She drew shapes on film to create a mask, which modulated the light received by photocells. That created audio-visual synchrony. Yet the monophonic sounds could be added to multitrack tapes to provide more texture, in which case there could also be non-synchronised sound/ image relationships.

[7] https://www.youtube.com/watch?v=0y-2shMhv0M
[8] http://vimeo.com/15919138
[9] https://www.youtube.com/watch?v=7cyHFT2abXE

Ryoji Ikeda's audio-visual performance *Test Pattern* (2008)[10] is a unequivocal example of the one-to-one approach. Ikeda uses a real–time computer program to convert audio signal patterns into tightly synchronised barcode patterns on screen. The velocity of the moving images is ultra–fast, some hundreds of frames per second.

Audio-visual artists working with digital 3D game engines often explore the one-to-one approach by taking advantage of 3D positional audio-effects, described in the previous section. For example, Tarik Barri's describes his audio-visual work *Versum* (since 2008)[11] as follows:

> The virtual world of *Versum* is seen and heard from the viewpoint of a moving virtual camera with virtual microphones attached. This camera, controlled in realtime by means of a joystick (or any other kind of controller) moves through space, similar to how first person shooter games work. Within this space, I place objects that can be both seen and heard, and like in reality, the closer the camera is to them, the louder you hear them. [http://tarikbarri.nl/projects/versum accessed August 12, 2016]

The structural approach can equally be illustrated with many examples. Thomas Wilfred's *Clavilux* (1919)[12] was a visual instrument that made use of multiple projectors, reflectors and coloured slides. It was mostly performed in silence. A significant exception was his performance in conjunction with the Philadelphia Orchestra, where Wilfred was careful to note that his aim was to create a specific atmosphere around the music, as opposed to following it note-by-note [Brougher et al. 2005].

Oskar Fischinger's *Lumigraph* (1940s)[13] was also a visual instrument, allowing for synchronised and non-synchronised audio-visual relationships. It produced imagery by pressing against a rubberised screen, protruding into a narrow beam of coloured light. Two people were required: one to manipulate the screen to create imagery, and a second to change the colours of the lights. Fishinger's inclination to the structural approach is reflected in *Motion Painting No.1* (1947)[14], an abstract animation accompanied with music by Johan Sebastian Bach. It is a film of a painting (oil on acrylic glass), and Fischinger filmed each brushstroke over the course of nine months. There are some moments of synchronisation, and other moments when the music is far more complex than the visual

[10] short excerpt at https://www.youtube.com/watch?v=GPhAvyrZ28o ; description at http://www.ryojiikeda.com/project/testpattern/
[11] http://tarikbarri.nl/projects/versum
[12] https://www.youtube.com/watch?v=ojVX8FWYc4g
[13] https://vimeo.com/13787280
[14] https://new.vk.com/video2804234_168196055

dynamics.

Tony Conrad's film *The Flicker* (1966)[15] is a clear example of the structural approach (and also an example of the structural film movement). While the music plays continuously, the image consists of only five different frames: a warning frame, two title frames, a black frame, and a white frame. Light and dark frames alternate with a rhythm that can produce after-images, seeing spots, and similar phenomena[16]. Conrad devoted himself to an intensive study of the physiology of the nervous system. He wrote: "I was interested in whether there might be combination-frequency effects that would occur with flicker, analogous to the combination-tone effects that are responsible for consonance in musical sound." [1996][17]

Steina and Woody Vasulka's *Heraldic View* (1974) [18], where analogue oscillators generate a visual pattern in a video, is another example of the structural approach. The sound and the generated visual pattern are synchronised, but there is an additional visual pattern, moving horizontally, independent from sound.

Phill Niblock's film series *Movements of People Working* (since 1973)[19] also incorporates the structural approach. These works create audio-visual relationships between the repetitive movements of manual labour and the surreptitious nuances embedded in massive drones of sound. There is no technological connection between sound and image; Niblock captures the film footage with a single shot, and creates related minimalistic scores for the music.

The work of computer animation pioneer John Whitney [Whitney 1980] is also a well-known example of the structural approach. Whitney used Pythagorean ratios to create a functional harmony of tension and resolution, with analogue and digital computers[20]. He used musical terms to describe how simultaneous sounds and images functioned interdependently (harmony) and evolved rhythm and contour (counterpoint). Whitney described the audio-visual relationship to oscillate between tension and release, and that same expression has been used to describe recent audio-visual performance works where the performers use computers, but no technological connection between sounds and images. An example is the work of Jentzsch and Ishii [2010]. The authors create the audio-visual relationship in

[15] https://www.youtube.com/watch?v=ZJbqnztjkbs
[16] http://www.medienkunstnetz.de/themes/overview_of_media_art/perception/5/
[17] Tony Conrad about *The Flicker*, http://flicker75.blogspot.pt/2008/01/flicker.html
[18] http://www.fondation-langlois.org/html/e/page.php?NumPage=481
[19] DVD published by Microcinema, 2003
[20] see for example *Arabesque*, 1975: https://www.youtube.com/watch?v=w7h0ppnUQhE

real-time, while exploring what they call metaphorical correspondences between volume and size, colour and pitch, timbre and brightness. Their approach is structural as well, because some sonic and visual events are synchronised and others are not.

Amongst those audio-visual works that explore the structural approach, many seek to emphasise the transformative power of technological mediation. For example, Gary Hill's early work was concerned with how electronic media transform linguistics. His film *Black White Text*[21] (1980) explores a relationship between geometric black and white figures and human voice. At the beginning, sound and image appear detached: the pace of the words and the visual changes is similar, but the rhythm is different, and there is hardly any synchrony. As spoken words accumulate, the sound becomes continuous, and language becomes progressively abstracted from any meanings beyond the sound itself. Another, recent example can be found in live coding, which emphasises technological mediation not solely because we see the code, but also because the visible code text will never synchronise with the generated sounds and/ or images: it must be typed before.

2.3.2 Music with Digital 3D Environments

Digital 3D engines can be used to create a sense of visual depth, and spatialise digital sounds with respect to the 3D scene. Beginning in the late 1990s, artists, musicians and programmers started exploring these technologies in audio-visual performance. A notable example is the *Global Visual Music Project* [22], developed by Vibeke Sorensen, Miller Puckette and Rand Steiger in 1997. It used Pure Data, the visual programming language developed by Puckette. The authors developed a system for audio-visual performance incorporating digital video, 3D graphics, acoustic musical instruments, and computer music strategies. This resulted in performances that took place in spaces networked through the Internet. The authors' primary concern was to achieve what Sorensen describes as an "abstraction of connection". Instead of making direct correspondences between pitches and colours, for example, they aimed at making connections that were clearly perceptible without being purely metaphorical. Rather than present direct visualisations of sounds or sonifications of visual elements, they connected different sites, and elements within a site, in a more semantic way. The program note for *Lemma 2* (1999) included the following note:

> Vanessa's cowbell in Oregon might sound as a similar cowbell in New York, but it might instead appear as a tomtom or as middle C on a computer-controlled piano. Moreover, computer graphics are shown at both sites which can respond in many

[21] https://www.youtube.com/watch?v=bg1O3NcPwBg
[22] http://www.visualmusic.org/

different ways to musical gestures at either location.[23]

My practical work also aims at creating non-metaphorical connections between sound and image, while binding a physical and a digital 3D space. Another common aspect is the fact that Sorensen and her colleagues intended their system to be played by specific musicians, as is the case with my work.

This was not the intent of many artists and researchers interested in music with digital 3D environments, who explored links between video game technologies and music so as to convey the audience's networked, participatory interaction with the system.

Shown for the first time in 1999, John Klima's *Glasbead* is a multi-user collaborative musical interface, consisting of a rotating spherical structure with stems that resemble hammers and bells [Mirapaul 2000]. Sound files can be imported into the bells, and triggered by flinging the hammers into the bells. Dealing with the concepts of multi-user environments, gaming and file-sharing, *Glasbead* allowed up to twenty players to remotely 'jam' with each other. Christiane Paul describes this piece as "an instrument and 'toy' that allows users to import sound files and create a myriad of soundscapes" [Paul 2008]. As an open-ended system, the work also matches Tanaka's definition of NIME [2009].

Wozniewski et al. developed a system where the user's body can be modelled within a digital 3D world[24] [2006a, 2006b, 2006c]. This system has been mentioned in chapter 2.3.3, which reviewed theories about the sense of spatial presence; in the authors' words, the work aims to superimpose the physical and digital space. The sound processing occurs at various locations in the virtual 3D space, and the sending of sound signals from one area to another is based on physical models of sound propagation. Initially, the system was intended for the authors' performance, but the subsequent versions were developed for participatory, networked interaction. The system is controlled by regular physical activity such as moving, turning, pointing, or grabbing.

Other works draw upon the most popular player paradigm in video gaming: First Person Shooters (FPS). Mick Grierson created a "3D first-person composition and improvisation system" designed "to look and behave like contemporary first-person computer games" [Grierson 2007]. The user can create, adapt and combine elements with varying physical attributes to produce musical structures. The system does not exhibit pre-determined

[23] http://www.visualmusic.org/gvm/lemma2.htm
[24] http://wikibin.org/articles/audioscape.html

constraints on the environment, object properties and interactions. It responds to the user behaviour through adaptive algorithms.

Also, Robert Hamilton describes a modification of a game engine in a paper titled “q3osc: or how i learned to stop worrying and love the <bomb> game” [Hamilton 2008a]. The engine outputs sound-objects for real-time networked performance and spatialisation within a multi-channel audio environment. The weapon projectiles are associated with sounds triggered on collisions with the environment. In another paper, Hamilton describes an “interactive multi-channel multi-user networked system for real-time composition” [Hamilton 2008b]. It makes use of navigation and sound spatialisation, by superimposing the virtual environment and the performance room. The software draws inspiration and structure from the topography of the virtual environment so that the interactor retains a level of control over the musical whole. It builds upon this visual and musical structure by leveraging the inherently flexible and indeterminate system of player motions and actions.

Furthermore, Florent Berthaut et al. developed *Couacs*, “a collaborative multiprocess instrument”, which led the authors to discuss the use of FPS for musical interaction [2011]. The work is a practical exploration of the following questions: “How can we use some game actions for musical control without disturbing other game actions not connected to sound, and vice-versa? Will gamers/musicians try to learn how the instrument works and how they can produce specific musical results, or will they only play without paying attention to the generated music? Should these instruments have a goal like a video game or not?” [2011]. Indeed, the definition of New Interfaces for Musical Experssion – NIME - embraces musical instruments as well as musical games.

The practical work of the thesis seeks to stress a difference between my personal creative concerns and the principles governing digital tools intended for video games - or participatory interaction in general.

The work of composer and cognitive researcher Jeff Pressing is important to this thesis in many ways. For a start, he identified several types of complexity relevant to music [Pressing 1987]. *Hierarchical complexity* refers to the existence of a structure across many levels. *Dynamic or adaptive complexity* refers to a rich range of behaviours over time, or an adaptation to unpredictable conditions, or a monitoring of results in relation to a reference source, or an anticipation of changes in oneself or the environment. *Information-based complexity* involves a target problem that requires a solution; since solutions are seldom exact, one actually seeks for an approximate solution, with a certain amount of tolerance.

Pressing also created a performance system for sound spatialisation in a digital 3D environment, problematising how the audio-visual relationship affects auditory perception [Pressing 1997]. He noticed:

> In virtual audio environments with high verisimilitude of simulation, as in simulated architectural acoustics, the signal carries a high information load with multidimensional perceptual implications [1997:8].

He proposed a related semantic characterisation of sounds that will be very useful in this thesis. According to this characterisation, a sound can be *expressive, informational,* or *environmental*. Pressing overlapped these typologies. Examples of expressive sound are all kinds of music and song. Examples of informational sounds are speech, alarms, and sonified data. Examples of environmental sounds include animal calls, wind sounds, and the noises of machinery.

Pressing observed that information-bearing sounds reinforce the visual identification of objects, and that this causes perception to prioritise the visual information over the aural. A primary aim in the present research is to understand whether and how that can be avoided.

2.3.3 The Artistic Debate on Sensory Dominance

In neurology, synaesthesia is a condition in which stimulation of one sensory or cognitive pathway leads to automatic, involuntary experiences in a second sensory or cognitive pathway [Cytowic 2002]. The term is often employed to describe audio-visual art/ performance, including the work of artists who do not suffer from synaesthesia. However, its use has long been criticised by researchers in audio-visual practice [LeGrice 1982].

The theoretical debates that surround audio-visual practice are greatly informed by Michel Chion's *Audio-Vision: Sound on Screen* [1994]. The book addresses the use of sound in narrative film, yet its theories equally apply to abstract sounds and images [e.g. Stevens 2009]. Those theories are not about sensory dominance, yet they are relevant to the debate on the subject. Chion developed the idea that the combination of sonic and visual elements generates effects and meanings that would not exist with sound or image alone: the product is greater than the sum. The combination of those elements generates a third audiovisual element. He coined the term *added value* to describe the expressive and informative value with which a sound affects an image, creating “the definite impression that this information or expression “naturally” comes from what is seen, and is already contained in the image itself” [1994:5]. Chion also coined the term *synchresis* to describe how sounds and images become associated as a result of synchronisation. Such association is a spontaneous psychophysical phenomenon, manifest in the reflexive merging of auditory and visual modalities [1994:63-64].

Chion stressed that sound in general is usually not the main focus of the audience's attention [1994:6]. Moreover, he wrote:

> Each audio element enters into simultaneous vertical relationship with narrative elements contained in the image (characters, actions) and visual elements of texture and setting. These relationships are much more direct and salient than any relations the audio element could have with other sounds. [Chion 1994:40]

He described three types of listening, or modes, which we can extend into the audio-visual domain. The first is *causal listening,* which "consists of listening to a sound in order to gather information about its cause (or source)" [1994:28]. Causal audio-visual perception is equivalent; it consists of listening to the sounds and viewing the images in order to gather information about their cause. The second mode is *semantic listening*, which "refers to a code or a language to interpret a message" [1994:28]. The same is applicable to semantic audio-visual perception. The third mode of listening is *reduced listening*. Chion provides perspective over the term 'reduced', by stating that hiding the source of sounds "intensifies causal listening in taking away the aid of sight" [1994:32]. In applying to the audio-visual domain, the thesis considers how 'reduced' might refer to stripping the perceptual experience of conclusive causes and meanings.

In using the term *reduced listening*, Chion makes a reference to composer Pierre Schaeffer, who coined the term in 1966. Schaeffer noted that we tend to treat sound as a vehicle to pursue other objects. He used this term to describe a listening mode that focuses on the traits of the sound itself, independent of its cause and meaning [1966:270]. Working with gramophones, Schaeffer used basic sound transformation techniques such as reversed playback, altered pitch/ speed and timbre to free his material from associated causes and meanings. He found that in this way, a sonic event could become amenable to the compositional treatment of rhythms and timbres, rather then merely an evocative symbol. The composer Francisco López, who performs in visual darkness, wrote about a particular sense of spatial presence in music: "Being 'inside' the sound (instead of listening to it) creates a strong feeling of immersion where your own body moves into the perceptive background" [2004].

The abstract filmmaker Stan Brakhage was also concerned with the surplus of meaning created by the audio-visual relationship. In contrast with Schaeffer and López, he wanted to highlight the musical quality of the images and their montage without the distraction imposed by sound. According to his wife Marilyn Brakhage, he felt that sonic music tended to dominate over the more subtle rhythms of vision [2008]. The large majority of Brakhage's films are silent. In his few later audio-visual movies, the primary principle was one of non-synchronisation, of breaking any direct connection between pictures and sounds. The audio-visual relationship was obviously not one-to-one; it was rather structural.

In sound art philosophy, Salomé Voeglin emphasises how seeing happens in what she calls a “meta-position”, away from the seen; according to her, this distance enables a detachment and objectivity that presents itself as truth [2010:12]. She promotes an auditory culture, emancipated from our dominantly visual culture. We can say that her perspective over vision is concerned with what Schaeffer and Chion called causal and semantic listening. Voeglin argues that by contrast, hearing is “full of doubt” because it does not offer a meta-position: “there is no place where I am not simultaneous with the heard” [2010:12]. We can say that “full of doubt” means inconclusive causes and meanings, regardless of whether ‘hiding the source of sounds’ intensifies causal listening or not.

The graphic performer Meghan Stevens is also concerned with visual dominance. She presented a theory about how music could remain dominant, or at least maintain the same level of dominance in audio-visual performance [2009]. She assessed that sensory dominance depends on what she called the characteristics of simultaneous sounds and images: narrative, synchronisation, rhythm, tempo-pace, meaning, emotion, structure, continuity, and genre-style. According to her, visual dominance becomes overwhelming when there is a conflict between sound and image [2009:32]. She also assessed that full congruency normally causes attention to focus on the composite rather than on the individual media [2009:21]. *Partial congruency* would permit the focus to remain on the music. Stevens supported the idea with certain experiments conducted by Tom Grimes, which involved different versions of TV news stories; their results indicated that a moderate match of sound and image leads to reduced visual attention [Grimes 1990:15]. Stevens noted that in audio-visual performance with abstracting sounds and images the decision of what is congruent would vary from one person to another, but several factors would create similar associations of music and image between people: our cultural and educational background, learned associations, and the context in which the music and image are presented. Stevens safeguarded that her theories “can rightly be criticized as being created from limited evidence” [2009:3]. Indeed, those theories rely on sonic and visual shapes, and involve too many subjective variables.

In fact, the question of how the audio-visual may favour the aural or the visual depends on how perception prioritises sensory information. Hence, it is useful to look at psychology and neuroscience so as to draw a scientific frame of evidence.

2.4 Perception, Structure and Complexity

This section looks at how perception prioritises information, establishing the scientific foundations of the perceptual approach developed in this research.

2.4.1 Perception

Perception is a process of multisensory synthesis; the same principles govern across multiple sensory modalities, all being mutually influential despite our possible unawareness of such interactions [Calvert et al. 2004]. The primary aim of the brain is to use the information derived from the various senses in order to detect, perceive and respond efficiently to objects and events. We need to survive.

From the early 20th century to the present, Gestaltist psychologists have described how we organise the perceptual field in the simplest and clearest way possible, deriving the meaning of the parts from the meaning of the whole. The principles of this perceptual organisation have been studied in the aural and the visual domains. Table 1 summarises these principles in their most basic form (please see next page). Research on their subtleties could proceed with A.S. Bregman [Bregman 1990], J. Tenney [Tenney 1980], B. Snyder [Snyder 2001] and F. Lerdahl & R. Jackendoff [Lerdahl & Jackendoff 1983] in the aural domain; M. Wertheimer [Wertheimer 1938], E. Rubin [Rubin 1921], K. Koffka [Koffka 1935] and S. Palmer [Palmer 1999] in the visual domain.

At any time, we are presented with a massive amount of stimuli. According to many researchers [e.g. Snyder 2001, Knudsen 2007], incoming information transits from *sensory memory*[25] into *short-term memory*, and subsequently to *long-term memory* from where it is constantly retrieved. Whilst *long-term memory* indefinitely stores a seemingly unlimited amount of information, the rapid decay of *short-term memory* submits the stimuli to strong competition. The term *working memory* refers to the structures and processes used for temporarily storing and manipulating information.

[25] According to these researchers, the sensory memories act as buffers for stimuli received through the senses, and each sensory channel has its sensory memory: *iconic memory* for visual stimuli, *echoic memory* for aural stimuli and *haptic memory* for touch.

Principle of *invariance*: to discern cohesiveness in a changeable form	
A visual object is recognised independently from rotation, translation, scale, elastic deformations, lighting and component features.	An auditory stream is perceived as an unit in spite of its changes over time [Bregman 1990].
Principle of *figure/ background*: to segregate a form from the continuum	
Visual figures are perceived as separate from their background [Rubin 1921, Palmer 1999].	Discerning a sound implies segregating the sound from the soundscape.
Principle of *proximity*: to group closely located elements	
Visual elements that are closer together are perceived as a coherent object [Wertheimer 1938].	Sounds are grouped as a single event when adequately proximal in time [Bregman 1990].
Principle of *similarity:* to group similar elements	
Visual elements that look similar are grouped as part of the same form [Wertheimer 1938].	Simultaneous sounds with different spectrums are grouped when exhibiting the same fundamental tone or when their frequency components exhibit similar onset time [Bregman 1990].
Principle of *common fate*: to group elements that change or move together	
Visual shapes with the same orientation are grouped as a unit [Wertheimer 1938].	Sounds that change together are grouped as a unit [Bregman 1990].
Principle of *good continuation*: to group of elements that follow a consistent, lawful direction	
Visual shapes seem inextricably related when a parameter changes progressively (e.g brightness, size, location) [Wertheimer 1938].	Sounds seem inextricably related when a parameter changes progressively (e.g pitch, loudness, interval).
Principle of *closure:* To enclose a form by overlooking any gaps in its cohesion	
A visual space is enclosed as we mentally complete its contour, or ignore any gaps in the figure [Wertheimer 1938].	A gap or deviation in a linear sound sequence does not impede us to perceive that sequence. Moreover, musical phrases are tied together, relating at higher level in perceptual organisation [Snyder 2001].

Table 1: Gestaltist principles of perceptual organisation

Bob Snyder (composer and cognitive researcher) describes how working memory can handle large amounts of information simultaneously, through cues. Perception *chunks* sensory information continuously throughout information processing, from low-level stages (e.g. a sound includes a multitude of sound frequencies) to high-level stages (e.g. identifying a sound involves retrieving memories). According to him, “the focus of conscious awareness could be thought of as being at the “front edge” of short-term memory” [Snyder 2001:50]. This “front edge” has an even smaller capacity than short-term memory. Snyder estimates that it holds three pieces of information at the most; Cowan estimates that it holds four [2001]; Gilakjani estimates that it holds between three and five [2012]. In any case, each of those few information pieces embeds a hierarchy of semi-conscious and unconscious information.

Snyder describes how categorising the sensory information enables us to operate based on assumptions. Events activate parts of long-term memory that have been previously activated by similar events. Snyder calls these long-term memories *conceptual categories*. Among these long-term memories, some become highly activated and conscious, whilst others – which he calls *semiactivated memories* - remain unconscious, forming expectations. When new situations counteract these expectations, they require new combinations of cognitive processing. Since this cognitive processing depends greatly on attention, attention is constantly being drawn to the novel aspects of situations.

Novel aspects of situations attract attention, but only when the brain prioritises their significance over its tendency to simplify the perceptual field. In fact, as put by the media researcher Herbert Zettl, "(o)ur mental operating system encourages a considerable perceptual laziness that shields us from input overload. (…) We often see and hear only those details of an experience that fit our prejudicial image of what the event should be and ignore the ones that interfere with that image" [Zettl 1998]. We should add that our mental operating system can be extremely active while simplifying the sensory information. There is no structured thought without abstraction, and there is no abstraction without perceptual simplification. Also, sometimes we pop back and forth between multiple interpretations, as shown with the Gestaltist notion of *multistability* [Rubin 1921]. According to the philosopher Alva Noë, perceptual experience acquires content thanks to our practical bodily knowledge: perception is "an activity of exploring the environment drawing on knowledge of sensorimotor dependencies and thought" [Noë 2004:228]. Indeed, one can consciously experience sensorial complexity while the brain seeks to form abstractions of that same complexity.

We can enjoy the process in music. Snyder describes how music plays with Gestalts of *closure* – with our psychological tendency to complete a shape despite any existing gaps [2001:59-60]. Closure gives us the sense of cohesion in a musical shape. Closure ties musical events together, relating them at a higher level of information processing. Partial closure occurs when some parameters of musical dimension fulfil expectations, while others do not. These incompletely closed musical phrases create expectations of eventual closure. Any more completely closed musical phrase appearing subsequently will then not only close itself, but also close prior, less thoroughly closed musical phrases.

Snyder notes that the most definite musical expectations of closure are created by a tradition of rules of learned patterns signifying that closure. Furthermore, closure derives from the relation between other Gestaltist principles (Table 1). Snyder draws from A.S. Bregman, who established a theoretical framework for discussing the process of auditory grouping and segregation [Bregman 1990]. Bregman found that the gathering of events that follow one another in time (*sequential integration*) happens if the sounds are sufficiently proximal in time and/ or pitch, or due to similarity in timbre or in loudness. The closer in time the events are, the more proximal in pitch they have to be in order to aggregate in a single stream. He also investigated how we integrate simultaneous sounds with different spectral content (*spectral integration*). It happens through *harmonic grouping* (components with the same fundamental pitch likely come from the same source), *frequency onset grouping* (frequency components with similar onset time likely come from the same source), *similarity of temporal evolution grouping*, and *"old-plus-new-heuristic" grouping* (perceptual continuation of an old sound at the presentation of a more complex sound). Indeed, every "musical tradition" must ground on common human experience, involving common psychophysical aspects.

2.4.2 Multisensory Integration

The question here is how we can enjoy complex sonic constructions in audio-visual performance, given the stimuli competition to access conscious awareness, and "the forging of an immediate and necessary relationship between something one sees and something one hears" [Chion 1994:5]. *Multisensory integration* is a crucial topic - the term refers to the set of processes by which information arriving from the individual senses interacts and influences processing in other sensory modalities, conveying a unified experience of multisensory events. Whilst the Gestalt psychology describes phenomena that we can be consciously aware of, multisensory integration happens automatically, at a lower level in information processing. Pursued by teams of experimental psychologists and neuroscientists, the research in multisensory integration studies perception in terms of feature extraction, which precedes conscious experience.

As we process the sensory information, divergences across the sensory modalities can produce phenomena known as multisensory illusions. For example, the *ventriloquist effect* is a phenomenon in which a sound is perceived to occur at or towards the location of a spatially disparate visual stimulus that occurs at the same time [Pick et al. 1969][26]. The *sound-induced double-flash Illusion* refers to how a single flash of light, presented

[26] Chion describes this perceptual phenomenon as a necessary result of synchresis.

concurrently with a train of various short tone pips, is perceived as two or more flashes [Shams et al. 2002]. The *McGurk effect* occurs when speech sounds do not match the sight of simultaneously seen lip movements, and this leads to a perception of a phoneme that is different from both the auditory and visual inputs [McGurk 1976]. Furthermore, the perceived duration of auditory events can be shortened or lengthened by conflicting visual information [Kobayashi et al. 2007, Wassenhove et al. 2008]. In audio-visual works, the phenomenon can be explored when at least a part of the sonic and visual events are not in synchrony.

Researchers in psychology propose that visual stimuli tend to dominate in the processing of spatial characteristics (as with the ventriloquist effect) and auditory stimuli tend to dominate in the processing of temporal characteristics (as with the sound-induced double-flash illusion) [Welch and Warren 1980]. Chion noted that audition supports vision:

> The eye perceives more slowly because it has more to do all at once; it must explore in space as well as following along in time (…) Why don't the myriad rapid visual movements in kung fu or special effect movies create a confusing impression? The answer is that they are "spotted" by rapid auditory punctuation. [Chion 1994:11]

In other words, "what we hear is what we haven't had time to see" [Chion 1994:61].

Audition may dominate in the processing of temporal characteristics, and be crucial to visual information processing. Yet, the timings and intervals of a sonic construction may not be experienced when that sonic construction is coupled with moving images. The visual tends to obfuscate the aural. Timings and intervals are formed through relations between sonic events, and many of these events (the most subtle ones in particular) may not reach conscious awareness.

In many scientific experiments where participants were presented with random auditory, visual and audio-visual stimuli in a rapid discrimination task, they failed to respond to the auditory component significantly more often than to the visual - the phenomenon is known as the *Colavita visual dominance effect* [Colavita 1974]. The Colavita effect might reflect a response bias, rather than an attenuation of auditory encoding due to the visual input [Spence 2009]. In other words, visual dominance might happen at a level of conceptualisation, rather than at a detection level. In fact, recent experiments conducted by Robinson et al. show that auditory input slows down the visual response, and not the inverse [2015]. Whilst audition dominates with infants, adults strategically bias their responses in favour of visual input, possibly to compensate for the poor alerting abilities of visual stimuli.

Conceptualising causation seems to have great influence in visual dominance. Since Pierre Schaeffer [1966], acousmatic composers argue that sounds must be detached from their visual origination to be fully experienced. Jeff Pressing also noted that, in digital 3D environments, "where a direction of causation is discernible, it operates from visual event to sound event and not the reverse. Nevertheless, sometimes the sound for a new cinematic scene can precede the visual image" [1997:10]. Indeed, 3D animators often place the sound of a footstep slightly before the foot actually hits the ground.

Visual stimuli tend to dominate in conscious awareness, but that can also be avoided:

> The visual channel is sampled before, or possibly more frequently than the auditory channel. However, because this difference in sampling rate (or bias) may be attentional in nature, it can be manipulated by focusing attention on one sensory modality or another. [Sinnet et al. 2007]

This conclusion is a steppingstone to the present investigation.

2.4.3 Attention

Given the central role of attention in sensory dominance, attention dynamics are of major importance here. Attention is *exogenous* or *bottom-up* when automatically driven through stimuli, and *endogenous* or *top-down* when under individual control [Knudsen 2007]. Typically, attention is drawn automatically to events that are infrequent in time or in space – stimuli that break perceptual continuity, as happens when a stimulus appears or disappears in sudden manner. These events are of high biological relevance, making for the nervous system to respond in strong and automatic ways. Conversely, attention is under greater individual control when there are no salient changes.

Time affects the interplay of endogenous and exogenous attention, but not in a linear way. A long period of consistent stimuli makes an inconsistent change more infrequent than a short period. Potentially, that change becomes more salient. However, consistency over time also allows for one to deviate attention, in which case the change might not be salient enough to break perceptual continuity.

Attention determines whether an event possesses greater or lesser resolution in the perceptual flux. E. Knudsen, a researcher in the field explains:

> At any point in time, the information that gains access to working memory is selected by a competitive process (...) Signal strength reflects the combined effects

> of the quality of the encoded information, top-down bias signals, and bottom-up salience filters. [2007:58]

Whether attention is automatic or deliberate, it causes working memory to optimise the resolution of information concerning its target. This occurs when the sensory organs are directed toward the target, and/ or when the sensitivity of neural circuits is modulated accordingly. Working memory then improves the quality of related information processing in all domains: sensory, motor, internal state, and memory.

This indicates the challenge of circumventing visual dominance in audio-visual performance. We see a maximum of 180º, but we can direct the eyes toward a target. We hear 360º, but without physically directing the ears. Even when visual stimuli do not attract automatic attention, as soon as we direct the eyes to the screen, by default, working memory optimises perceptual resolution according to the visual target. We can say that optimising perceptual resolution for the music depends on a mental decision, rather than a physical act.

A question is, do automatic attention and deliberate attention have equal impact in multisensory integration? It has been shown that endogenous (deliberate) attention has an influence when there is considerable competition between inputs to different modalities, and the attention target has to be selected deliberately [Talsma et al. 2010]. Scientific experiments occur in very controlled environments, thus one should be careful with extending certain conclusions to situations where many diverging stimuli are competing to reach conscious awareness.

That being said, a number of experiments have demonstrated how automatic attention influences multisensory integration, and deliberate attention does not. For example, the ventriloquist effect has been shown to largely reflect automatic sensory interactions, with little or no influence from deliberate attention [Bertelson et al. 2000]. Similarly, Van der Burg et al. showed that, given a sound and a specific target in a visual field, the target only pops out if changes in the sound and image signals are both synchronised and abrupt [Van der Burg et al. 2010]. In other words, if sound and image prompt automatic attention simultaneously.

That can happen in our two types of approaches to audio-visual performance and animation, which I called the one-to-one approach and the structural approach. In McLaren's *Dots*[27] and Ikeda's *Test Pattern*[28]*,* abrupt audio-visual changes synchronised

[27] http://vimeo.com/15919138

one-to-one cause visual targets to pop out each time there is a sound. In Vasulka's *Heraldic View*[29], where the oscillator sounds synchronise with equally abrupt visual changes, these pop out relatively to the visual pattern that moves independently from sound. In multisensory integration, the abrupt/ synchronised changes will always dominate over the other changes. Furthermore, automatic attention might not depend on synchrony because sudden visual changes also dominate in non-synchronised audio-visual relationships. Conrad's *The Flicker* [30] is a good example: it explores perceptual effects driven by the abrupt visual switches between light and dark, which dominate over the static image periods while the music plays continuously.

The thesis' approach to the question of sensory dominance is encouraged by the fact that attention can be manipulated so that vision does not dominate over audition. In summary, the questions now are: how can working memory optimise perceptual resolution according to the music, if the audience directs the eyes toward the screen? How does the stimuli panorama affect the interplay of automatic and deliberate attention? And how does attention condition how we perceive the stimuli panorama? We need an operational tool for comparing the strength of sound and image. It is very helpful to understand that attention is automatically driven to stimuli that are infrequent/ out of context, and that perceptual discontinuity increases neural activity [Knudsen 2007].

2.4.4 Audio-Visual Objecthood

Merilyn Boltz investigated how congruent and incongruent sound/ image pairings affect perceptual encoding and memory representation in film [2004]. The mood of a movie soundtrack can be congruent or incongruent with the mood of the visual action. Boltz observed that mood-congruent sound/ image pairings result in integrated perceptual encodings and representations, where vision dominates over audition. Conversely, incongruent pairings result in independent encodings and representations for each sensory modality.

This indicates that attention can be led to prioritise the relations amongst the sounds over the relations across sensory modalities. Music might often not evoke a specific, clearly identifiable mood, particularly outside the realm of narrative film. Boltz adopts the idea that the mood of music depends on the interplay of pitch, timing, and loudness. One could

---

[28] short excerpt at https://www.youtube.com/watch?v=GPhAvyrZ28o
[29] http://www.fondation-langlois.org/html/e/page.php?NumPage=481
[30] https://www.youtube.com/watch?v=ZJbqnztjkbs

argue that it depends greatly on internalised traditions, which in film lead people to associate those musical characteristics with specific moods. Nevertheless, when narrative film is excluded from the equation we can equally say that congruence leads to integrated perceptual encodings/ representations, and incongruence leads to separate encodings/ representations.

M. Schutz and M. Kubovy conducted a study about the perception of a percussive action, recorded on video [2009]. By manipulating the synchronisation between sound and image, they observed that synchrony does not inevitably lead to automatic sensory interactions. Such interactions depend on the strength of perceptual binding, which in turn depends on what Kubovy and Schutz call the *ecological fit* between auditory and visual information [2010]. For example, seeing the strike of a marimba fits naturally with a percussive sound, but not with a piano sound. Thus, when the sound and the image are desynchronised (up to 700ms), perception shortens the marimba sound, but not the piano sound [Schutz and Kubovy 2009]. Kubovy and Schutz explain that multisensory interactions consist of a reciprocal mapping between the aural and the visual information [2010]. The perceptual binding is not merely associative: the visual discounts the aural and the aural discounts the visual. The ecological fit is assessed through concepts, which they call *audiovisual objects*, stating: “even though we perceive the world and not percepts, we cannot dispense with mind-dependent concepts, and indeed entities” [2010:58].

Ecological fit means congruency, or plausibility. The marimba experiment stresses that our assessments about congruency are greatly based on concepts of causation, and this is also clear in Gestaltist psychology. Indeed, we are driven to assess the cause of stimuli, and that informs perceptual binding. Sensory prioritisation is greatly driven by the primary aim of the brain; we need to make sense of the environment in good time so as to survive. Hence, the meaning of the parts depends on the meaning of the whole.

In summary, we can apply Kubovy’s notion of audiovisual object to perceptual binding in general, from low level to high level in perceptual organisation. It is helpful to know that synchrony is not bound to skew one sense in favour of the other. The question then is whether, and how, the audio-visual relationship can produce a sense of causation without appearing fully congruent.

2.4.5 Intensity

To make sense of reality is to create some type of perceptual continuity. This applies to the auditory, the visual and the audio-visual domains, as shown with Gestaltist principles and multisensory integration processes.

The composer James Tenney investigated which "are the factors leading to the discovery of continuity" in music [1980]. He proposed that the cohesion and segregation of sounds is primarily determined by the Gestalts of *proximity* and *similarity*, i.e. the perceptual grouping of closely located and similar information. For Tenney, intensity would be among the secondary factors. He defined intensity as:

> (…) the tendency of an accented sound to be heard as the beginning of the [perceptual] grouping. The relative intensities of several concurrent elements in a clang (or several monophonic sequences in a polyphonic sequence) are also a determinant of textural focus. [1980:90]

Tenney spoke of "musical or subjective intensity"; generally, intensity could be associated with increases in loudness, pitch, harmonic dissonance, tonal brightness, speed or temporal density. In his view, one would possibly never be in a position to describe the factor of intensity in a "satisfactory way". The research of this thesis hopes to update the reader's opinion.

Snyder adopted Tenney's notion of intensity [Tenney 1980:33-41]. Both researchers define intensity as a physical property of stimuli, related with preceding, simultaneous and subsequent stimuli. Snyder summarised the notion as follows: intensity is "any change in the chain of stimuli that causes an increase in neural activity" [Snyder 2001:62]. Furthermore, he described *musical motion* as a continuous oscillation between points of high intensity (corresponding to motion or tension) and points of low intensity (corresponding to rest or release). He explained that points of high intensity indicate instability, i.e. lack of closure. Conversely, points of low intensity indicate stability, i.e. closure: musical structures are resolved when a sense of arriving at a "goal" has been achieved.

Although intensity is usually understood as a property of stimuli, it is good to keep in mind that the term cannot be detached from perceptual impact. That applies to both the auditory and the visual domain. Also, it is good to keep in mind that 'subjective intensity' (using Tenney's term) is important in the *discovery* of continuity in music.

2.5 Sonic Expression, Interaction and Flow

The idea of discovery is useful when looking at what musical expression means. Understanding how perception works is crucial to certain research in music, and it also provides useful perspective upon artistic discourses that are seemingly distant from science. The question of how an instrument can convey expression has several layers, and it requires considering different creative perspectives. This section reviews literature in interaction design, music and psychology, situating how different notions of musical expression might govern different types of interaction design.

2.5.1 Interaction

In the field of interaction design, William Gaver distinguished three types of mappings between data and its representation, applying them to sound [1986]: *symbolic* (arbitrary mapping), *metaphoric* (similarities between data and representation) and *nomic* (direct relationship between representation of the sound source and sound). Gaver coined the term *auditory icons* to describe nomic mappings, which are based on the way people listen to the world in their everyday lives. He stressed that these mappings can provide information about sources of data, and simultaneously make the interaction seem natural.

The interaction seems natural when the medium feels like a natural extension of the body. The term embodiment has many meanings and ramifications. We can look at how it has been used by Paul Dourish, a researcher in embodied human-computer interaction [Dourish 2004]. Dourish described how interfaces apply the study of those mechanisms through which we naturally experience the world, such that the user overlooks the medium/object and focuses on that which is being mediated. For example, as the dynamic navigation in a 3D video game reproduces vision and audition in the physical world, it also conveys the player's immersion in the game.

According to Dourish, embodiment means possessing and acting through the physical manifestation of the world, and embodied phenomena occur in real time and real space. The notion of physical presence is extended to include phenomena that may not be physical but occur in the real world: we engage with the non-physical world because things occur there. Furthermore, Frances Dyson pointed out that the notion of embodiment is frequently theorised upon as if it was exclusive to digital media, which makes the term slippery; in fact the auditory experience is itself embodiment, as sound cannot be touched, we hear 360º, and we cannot shut our ears [Dyson 2009]. Both Dourish' and Dyson's notions of embodiment are relevant to this thesis: whereas the former emphasises interaction, the latter emphasises perception.

2.5.2 Sonic Expression: "Chance", Intensity, Effort and Timing

Our ways of interacting with the world condition how we perceive the world, and vice-versa. In the 1950s John Cage proposed that uncontrolled features can be used as musical material[31], and many recent artists explore related compositional approaches [e.g. Vasulka 1996, Cascone 2000]. Cage explored the musical potential of noises and sounds with indeterminate pitch, employing principles of rhythmic organisation distinct from those of melody and harmony [e.g. see Vergo 2010]. He tossed I-Ching coins and Tarot to score many of his pieces, exploring chance as a compositional principle. His strong assimilation of eastern philosophies is well known, and these philosophies suggest that suppressing intention is required to permeate the unity and mutual interrelation of all things, which are inseparable parts of a cosmic whole. Cage studied Indian philosophy and music. When he asked what was the purpose of music in Indian philosophy, he was answered: "to sober the mind and thus make it susceptible to divine influences" [Cage 1961:158]. This 'sobering of the mind' seems in direct relation with that which Francisca Schroeder and Pedro Rebelo called the *performative layer* [2009]. They coined the term to describe how the performer's strategies in dealing with discontinuities, breakdowns and the unexpected reflect "a becoming-aware-of and awakening of unused abilities".

Snyder's notion of musical motion [2001] reflects the fact that counteracted expectations cause an increase in neural activity, and fulfilled expectations cause a decrease in neural activity. In the previous section we saw that neural activity reflects attention, and we saw that sudden discontinuities prompt automatic attention [Knudsen 2007]. Indeed, the relation between attention, continuity and discontinuity is often important in music analysis. For example in *Spectromorphology* (a field of musicology that studies the perceived morphological developments of sound spectra over time), Smalley describes how attention is drawn to the initiation and the termination of a sound if the sound has a sudden onset that is immediately terminated, or when the onset is extended by a resonance that decays towards termination [1986]. Conversely, when both the onset and the termination are graduated, attention is drawn to how the sound is maintained (ibid.).

[31] Marcel Duchamp's *Erratum Musical* (1913) anticipates for nearly forty years John Cage's thinking of how the laws of chance might apply to music. Duchamp created a musical *jeu d'esperit,* and Cage explored chance as a compositional principle [Vergo 2010].

So how can the interaction with a musical instrument convey an expressive interplay of continuities and discontinuities? Joel Ryan, who pioneered digital signal processing of acoustic instruments, wrote that it is interesting to make control as difficult as possible, because effort is closely related to musical expression [1991]. Andrew Johnston speaks of a *conversational* type of interaction with digital music instruments: “the musician allows the virtual instrument to ‘talk back’ (…) responsibility for shaping musical direction continually shifts between musician and virtual instrument” [2011:293]. Furthermore, Tanaka stresses the importance of volatility in expression, and he also distinguishes volatility from unpredictability [Sa et al. 2015].

In the preface interview of the ICLI 2014 proceedings, Ryan, MacPherson, Magnusson and Tanaka agreed on the importance of timing for musical expression, despite their very different digital music interfaces and performance composition methods [Sa et al. 2015]. Inclusively, the title of Ryan’s contribution to the first round table of the conference was *“Knowing When”.* In the subsequent interview I asked him to clarify this title, and he introduced the term *local time*:

> The fact is I know when. Before it happens, I know when a beat should come, I know after, when it didn't. This knowledge is not something you can necessarily explain in words. It is something you demonstrate in playing but also listening, in enjoying music. It is the knowledge of how to make time. The proof is that with practice you get there on time, again and again. (…) The time referred to here is not the objective, uniform time inferred by physics or fashioned by technology, but another, local time. It is not a supplement or embellishment nor is it a primitive or schematic time but the time we make, enacted time, dense and polyvalent, the most elaborate aspect of time in music. [Sa et al. 2015]

We can say that this enacted, musical time is simultaneously personal and universal: indeed, the audience is equally sensitive to its logics. In my view, it reflects an “awaken of unused abilities” (quoting [Schroeder and Rebelo 2009]); which in turn requires the musical interface to be effortful, to a certain extent. Effort can be considered relatively to the amount of cognitive processing required in the construction of musical time. The thesis will elaborate on this notion of musical expression, yet to parameterise interaction we need to consider different notions as well.

2.5.3 Flow

Repurposing video game technologies so as to convey idiosyncratic expression brings the question of how the notion of ‘play’ might differ in gaming and music. Interestingly, the term New Interfaces for Musical Expression (NIME) can embrace musical instruments and musical games. It indicates that the distinction may not be consensual. We can speak of musical expression in both cases, and yet my personal understanding of expression seems distant from gaming. So how can the notion 'flow' be understood in music and gaming, and how does the difference substantiate in interaction design?

Psychologist M. Csikszentmihalyi sought for a general theory about flow [1996, 2004], which became influential in game interaction [e.g. Koster 2004, Sweetser and Wyeth 2005]. According to him, flow entails: concentration, merging of action and awareness, loss of self-consciousness, transformation of time, balance between challenge and skills, clear goals, direct feedback, and sense of control. It is clear that musical flow requires the interaction to enable concentration: being aware only of what is relevant here and now. It also requires merging of action and awareness: action exceeds conscious mind. And, it implies transformation of time: the sense of how much time passes depends on what one is doing. Beyond these convergences, the other features of flow require clarification.

Csikszentmihalyi states that flow entails *clear goals* every step of the way, meaning that one always knows what needs to be done. To some extent this applies to both gaming and music, but we need to define which kind of knowledge we are speaking about. Media researchers find that "to induce optimal presence, the developer of a mediated experience has to include recognition of the specific purpose of the user" [Riva et al. 2015:90]. Indeed, a video game player needs clear goals to act strategically. Even in sandbox games such as *The Sims* [32], where there isn't any narrative goal, players have clear goals such as making a bigger house to host more inhabitants, or decorating characters to make them attractive for social interaction with other game players. In contrast, many musicians feel that musical actions can be clear and deliberate without their motivations being clearly conscious. That is also implicit in the notion of performative layer, which entails the wakening, extending and refining of unused skills [Schroeder and Rebelo 2009].

According to Csikszentmihalyi, another feature of flow is *loss of self-consciousness*:

> In everyday life we are always monitoring how we appear to other people; we are on the alert to defend ourselves from potential slights and anxious to make a favourable impression. Typically this awareness of self is a burden. In flow we are too involved with what we are doing to care about protecting the ego. [Csikszentmihalyi 2004]

Csikszentmihalyi speaks as a psychologist. A performer can argue that the sense of self also exists beyond any anxiety to make a "favourable impression". As a musician, when I am on stage indeed I am too involved with performative action to care about protecting the

[32] E.g http://www.thesims3.com

ego, and yet, I am highly aware of how I appear to the audience. It is easy to agree with Csikszentmihalyi when he writes that “the musician feels at one with the harmony of the cosmos” [2004:12-13]. It becomes more difficult when he states that such feeling only comes *after* an episode of flow is over, because “the nervous system has definite limits on how much information it can process at a given time” [2004:28].

Another feature from Csikszentmihalyi‘s notion of flow is *balance between challenge and skills.* We can say that such balance exists in both gaming and music, yet possibly at different levels. The game industry aims to rapidly engage the player, and difficulties are adjusted accordingly. Many designers seek methodologies for video games to adapt to different types of gamers, while keeping all of them engaged [e.g. Hunicke and Chapman 2004]. There are very different approaches in musical instrument design.

Sense of control is another feature from Csikszentmihalyi’s flow. Again, this might be applicable to both gaming and music, but possibly at different levels. In gaming, interfaces are required to behave in consistent, i.e. clearly perceivable ways. There are different opinions in music.

Csikszentmihalyi’s notion of flow also entails *direct feedback*, i.e. a sense of immediacy. 3D video games apply the study of human perception mechanisms so that the player embodies the interface, and his action feels unmediated [Dourish 2004, Riva et al. 2015]. The problem is more complicated in music. Music is construction of time, and direct feedback is an equivocal term when we speak of digital music instruments.

2.5.4 Flow in Music

Csikszentmihalyi’s model of flow might be compatible with some understandings of musical expression, but many aspects are incompatible with the understanding presented in this thesis. The first debatable feature in his notion of flow is ‘clear goals every step of the way’. As Zbigniew Karkowski wrote, music can "heighten consciousness" and "increase intensity of the mind" because “art communicates before it is understood” [1992].

The second debatable feature is ‘loss of self-awareness’. Musicians are often well aware of permeating a wider sensory complexity. We can say that a musician’s empathic bind with the audience grounds on common semiactivated memories (using the term introduced by Snyder [2001]): it grounds on relating previous concepts with common, human biophysical rhythms. While the musician handles this latent, technically unconscious information as creative material, she does not loose awareness of self: she rather expands that awareness.

The third debatable feature of flow is 'balance between challenge and skills'. Game designers seek to adapt the level of challenge to different types of gamers [e.g. Hunicke and Chapman 2004], and there is also a salient, related discourse in musical interface design. For example, Francois Pachet developed what he called *musical mirroring effects*, whereby construction, the level of challenge represented by the behaviour of the system always corresponds to the level of the user [Pachet 2004]. Another example are the *personal instruments* developed by Machover and the MIT Media Lab, which the authors describe as musical tools that enable everyone to participate directly in music-making, regardless of their background [Machover 2009].

Alternatively, one can defend that an instrument requires great investment in playing. This is often the case when the musician develops his own instrument. For Michael Waisvisz, changing the algorithms that constitute the sound engine meant learning a new instrument, involving the re-incorporation of the conceptual understanding of the engine's functionality into bodily memory [1999]. J. Cannon and S. Favilla also stress that creating a new instrument must be accompanied with developing new skills to play the instrument; one does not learn to play an acoustic instrument in weeks, and that should also not be expected with digital instruments [Cannon and Favilla 2012]. Game interfaces might also be effortful and require one's investment of time, but to my knowledge this is never the case with musical games.

The fourth debatable feature in Csikszentmihalyi's model of flow is 'sense of control'. There are different opinions in music about the performer's sense of control over the instrument/ system. Poepel and Overholt's survey concluded that an absolute control over the instrument and clearly understandable relationships are desirable in NIME design [2006]. In contrast, Tom Mudd argues that non-linear behaviours foster engagement with musical expression, as the musician focuses on the material properties of the interface itself [Mudd 2015]. In consonance with this idea, the preface interview of the ICLI 2014 proceedings shows that McPherson, Tanaka and McLean do not enjoy an absolute, deterministic control over the interface [Sa et al. 2015]. Indeed, many NIME designs aim for the interaction to be learnable and repeatable, but complexity and volatility should exceed the performers' absolute control. For example, Tanaka thinks that "to control everything deterministically is not very interesting, and wouldn't give life in music" [Sa et al. 2015]. And McPherson thinks that "complex sonic behaviours absolutely have a role in digital instruments, as they do in acoustic instruments (e.g. woodwind multiphonics, certain string articulations)" [Sa et al. 2015]. Furthermore, Mick Grierson puts complex interface behaviours into perspective. He finds that:

> We spend a great deal of time practicing complex behaviour. This virtuosity has a tendency to infect musical and sonic style in a negative way. I can think of very few instances when this has resulted in music that expresses anything other than 'look how great I am'… Complex spaces of interaction and behaviour are great, but it is finite, specific interactions and behaviour that carry meaning. These don't require complexity at all. [Sa et al. 2015]

In any case, we can say that the performative layer from [Scroeder and Rebelo 2009] entails a risk, because it entails relying on the performer's skills to create un-prescribed musical meanings upon unexpected events - which the performer and the audience perceive at the same time. There are many different types of risk, and creative perspectives. For example, Magnusson wrote that where the digital music instrument exhibits any chaotic or entropic behaviour, this is usually seen as failure in design, a bug in the code or loose wiring in the hardware [Magnusson 2009]. At the same time, those behaviours are often also regarded as creative material [Cascone 2000]. Furthermore, we may find that the unexpected must be clamped so as to rule out undesired outcomes. As Ryan points out, "each link between performer and computer has to be invented before anything can be played (…) These 'handles' are just as useful for the development or discovery of the piece as for the performance itself" [Ryan 1991:5].

The fifth debatable feature of flow is 'direct feedback'. The question is more complicated in music than gaming. 'Direct' implies control over timing, and the construction of time is at the heart of music. But the term is equivocal. With digital systems operating with EMG[33], we can say that actuation happens faster than with any acoustic instrument, because the biosignal is captured when the performer initiates the physical gesture (see Tanaka's contribution to [Sa et al. 2015]). With live coding systems, the term immediacy gets even more ambivalent: typing code is an embodied and time-based action, but there is no one-to-one relationship with the sonic results. Also, about stand-alone digital music compositions Ryan wrote:

> Building from representations alone loses the open empiricism of play, and its desire to go beyond itself. (…) Time in music derives from performative knowledge. Systems of representation are capable of rendering many parts of this, but rendered via rigid symbol systems for discursive thinking, which moves more slowly than music. [Sa et al. 2015]

[33] Electromyography (EMG) is an electrodiagnostic medicine technique for evaluating and recording the electrical activity produced by muscles.

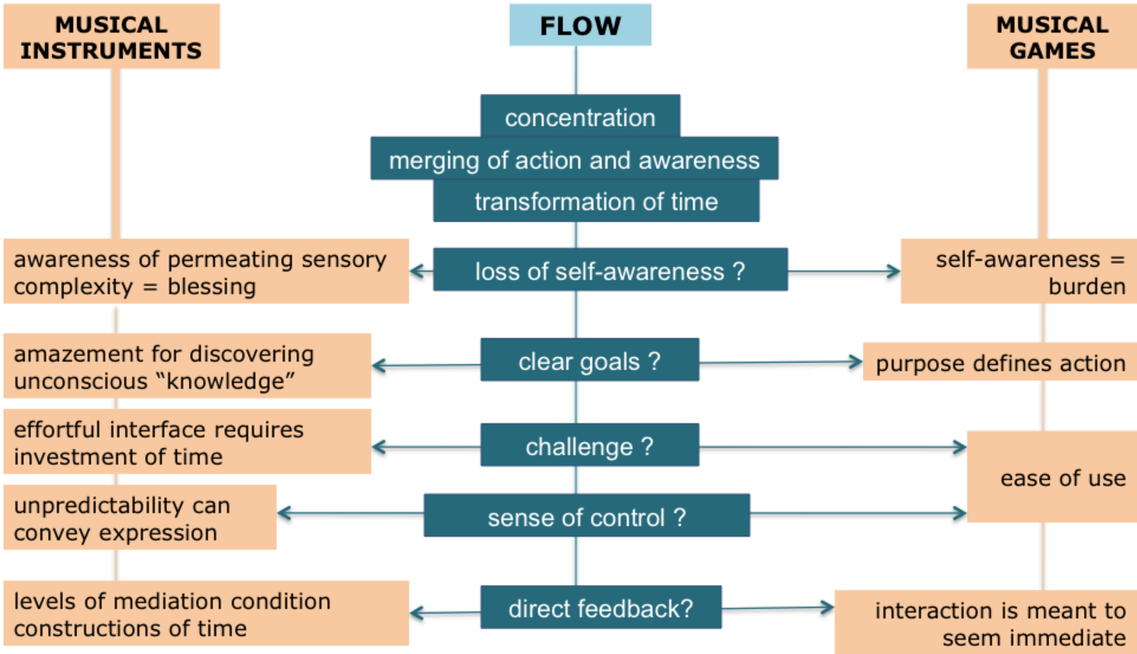


Figure 1: An arguable distinction between the notion of flow in music and gaming

Figure 1 summarises how my distinction between the notion of flow in musical instruments and games is reflected in interaction design. In both music and gaming, flow requires the interaction to enable concentration, merging of action and awareness, and transformation of time. But whilst a game player needs clear purposes, music has no purpose beyond perceptual experience itself - recalling Karkowski, art increases the mind's intensity because it communicates before it is understood [1992]. Whereas a game player looses self-consciousness as he focuses on the game, a performer is highly aware of her expanded state of perception, and of how the audience is being affected. Also, sense of control might be desired in both musical instruments and musical games, but at different levels and for different reasons. In gaming, interfaces are designed so that interaction feels as unmediated as possible. In contrast, with a musical instrument the performer focuses on the properties and non-linear behaviours of the interface, and the unexpected can convey expression.

Regardless of whether a musical device is called a musical instrument or a musical game, effortless musical interfaces indicate that the ideas governing sound organisation diverge from the notion of musical expression of this thesis. This makes effort a variable parameter in interaction design: other notions of sonic expression may require little effort. Effort also seems a variable parameter amongst those aesthetic approaches that fit the notion of expression of this thesis. This has at least two implications. The first is that there are many possible types of effort, motor and conceptual. And the second is that ultimately, all sorts of effort report to the challenge of music: making time, on time.

Effort is important to the notion of musical expression formulated in this thesis, but embodiment is equally important because it allows for one's natural interaction with the instrument/ system. The practical question then is to which extent, and how, one might want to reconcile these two aspects of interaction.

2.6 Discussion and Methods

Now that we have drawn the foundations of this research, I will discuss the literature reviewed in this chapter so as to justify the use of diverse, complementary research methods.

Regarding sound organisation, my practical question is how the combination of acoustic and digital components can potentiate sonic expression. Magnusson's thinking about digital mediation vs. acoustic immediacy [2009] is useful to situate my approach to interaction design, which involves the performer controlling the instrument without fully predicting the outcome. This approach does not coincide with that which Johnston calls conversational interaction [2011], because in my work the musician and the instrument "talk" at the same time. The instrument is simultaneously an extension of the body and a means of destabilisation. That which makes a sonic event "right" or "wrong" is very subtle. A digital sound arising unexpectedly could feel "wrong" within the musical logic, yet that logic can shift such that the event becomes gloriously "right". Whereas the unexpected event creates tension, the immediacy of an acoustic instrument allows me to incorporate the event in the musical logics, by shifting the musical direction with semi-conscious precision, boldly or surreptitiously.

My understanding of sonic expression relates Cage's "chance" [1961], Ryan's "local time" [2015] and "effort" [1991], and Schroeder & Rebelo's "performative layer" [2009]. When "chance" acquires musical sense, we can feel that music permeates a major reality. In fact, today one can speak of a major order in scientific terms. Chaos Theory, the existence of an order underlying apparently random data and/ processes, embraces Karl Popper's *causal chance*: "It was only the incompleteness of our knowledge which gave rise to this kind of chance" [Popper 1982:125]. Indeed, music can make us realise that we know more than we think we know; Snyder's notion of "semiactivated memories" [2001], which refers to semi-conscious perceptions, will help us to picture how this can be. Furthermore, the notion of the performative layer must imply effort and local time, because the "awakening of unused abilities" (quoting Schroeder) happens as the performer deals with "discontinuities, breakdowns and the unexpected". I will attempt to describe how we can be driven to permeate sensory complexity beyond that which we are capable to rationalise; this aim will lead me to bridge psychology, neuroscience and philosophy. Looking at human-computer interaction, music and NIME design, I will further argue that in many musical practices, the interaction with the instrument entails not solely embodiment, but also a separation that is reflected in expressive effort.

We can parameterise sonic expression in terms of interaction and perceptual dynamics. My personal understanding of expression requires an effortful interaction with the instrument, and practice. The term "effort" seems preferable to the term "virtuosity", which might not account for the creative potential of the unexpected, due to its origin in classic music tradition. Clarifying how different notions of flow substantiate in interaction design was advantageous, because it would not make sense to parameterise effort unless some notions of expression required much less effort, and other notions required much more. Furthermore, regarding perceptual dynamics, Snyder's notion of musical motion [2001] is useful to describe how music provokes increases and decreases in neural activity; particularly because he puts these increases/ decreases in correspondence with different levels of intensity. We have seen that the amount of neural activity reflects attention dynamics [Knudson 2007]. This means that we can draw a direct relation between *intensity and attention*, and exploit an operational tool to analyse perceptual motion, be it sonic, visual or audio-visual.

Regarding the visual dynamics and the audio-visual relationship, my practical question is how the image can function as a stage scene that reacts based on the sound, but without distracting the audience's attention from the subtle nuances of the sonic construction. The literature review of this chapter has situated the problem of sensory dominance. Whilst Schaeffer [1966], Pressing [1997] and Stevens [2009] argued that vision tends to dominate over audition, Brakhage assessed the inverse [Merylin Brakhage 2012]. Indeed, sensory dominance goes both ways. In any case, the artistic debate on the subject has focused on audio-visual congruency, perceived causation and synchrony. To understand the premises of perceptual binding, the chapter bridged the seminal work of Chion [1994] and Pressing [1997] with recent research in psychology and neuroscience. We learnt that perception is conditioned by the primary aim of the brain [Calvert et al. 2004] and by the limited capacity of short-term memory [e.g. Snyder 2001]. This seems a logical explanation for why perception discounts sensory information so as to map stimuli to previous concepts - which Kubovy and Schutz call audiovisual objects [2010]. Furthermore, attention influences multisensory integration, and it can be manipulated so that vision does not dominate over audition in conscious awareness [Sinnett et al. 2007].

This means that we need to consider which parameters condition attention dynamics. It is interesting that the artistic debate about sensory dominance never focused on attention itself, particularly because attention is considered in music analysis [e.g. Smalley 1986]. Neuroscientific research shows that sudden discontinuities prompt automatic attention, and that attention increases perceptual resolution regardless of whether it is automatically driven by stimuli, or under individual control [Knudsen 2007]. Once we have related intensity with attention so as to analyse perceptual motion, we can use this operational tool to compare the strength of sound and image. Furthermore, we must also consider how congruent and incongruent audio-visual relationships condition attention. Visual dominance seems associated with conceptualisation processes [Spence 2009]. The researches of Pressing [1997] and Boltz [2004] indicate that congruency must be avoided if we want attention to prioritise the relations between the sounds over the relations across sensory modalities. Avoiding absolute congruency does not suffice: we must take into account that perception can ignore diverging sensory information so as to form congruent representations - it happens with multisensory illusions as well as Gestaltist principles. Kubovy and Schutz observed that the strength of perceptual binding depends on previous concepts [2010], and I will draw from their notion of audiovisual objects so as to distinguish how conclusive and inconclusive concepts of causation lead to different types of binding. A study on audio-visual mapping and perceived causation will demonstrate the difference: whilst clearly perceivable cause and effect relationships lead perception to segregate fitting and non-fitting information, confounded relationships loosen the perceptual hierarchy, which can make for one's sense of causation extend to the audio-visual relationship as a whole.

Regarding the audience's sense of spatial presence, my practical question is how the work can intertwine the physical and the digital 3D space. Emmerson's term 'performative arena' is useful to describe the physical and psychological space of the work as a whole [2007]. I will consider spatial presence as a cognitive feeling [Schubert 2009], which depends on attentional processes [Draper et al. 1998] and perceptual cues [e.g. Dourish 2004]. We have seen that according to Sacau et al. [2008], the sense of spatial presence in film depends primarily on the active suspension of disbelief, and in media, on the properties of stimuli. But in fact, spatial presence in film also depends on properties of stimuli. Both film and digital 3D environments with multi-channel audio explore the perceptual mechanisms through which we naturally experience the world. This conveys a sense of spatial presence in the "virtual" space beyond the screen [e.g. Dourish 2004]. For example, a visible sound source on the left of the screen is audible through the left speaker. Moreover, we expect to continue hearing that sound source if the camera moves to the right and the sound source disappears from the screen - as if we had just deviated our eyes from a physical sound source, in the physical world. Chion's notion of superfield denotes that perception does not need highly congruent audio-visual relationships to create a sense of continuity [1994]. Film is often playful with perceptual cues, in ways that twist the viewer's sense of congruency. Conversely, the dominant paradigm with digital 3D environments is to strive for realistic simulation. That paradigm corresponds to that which Kubovy and Schutz called a high ecological fit [2010]. It is inconvenient here, because absolute congruency is problematic for sonic experience. Extrapolating from the abovementioned study about audio-visual mappings and perceived causation, the thesis will also describe how my practical work addresses the audio-visual relationship in space.

But spatial presence and perceptual cues must not be solely considered with respect to sensory dominance. My musical forms, which combine acoustic and digital sounds, can create environmental soundscapes, and also conduct attention to the performer's expressiveness. At the same time, the dynamic view upon a digital 3D world - a deserted landscape that morphs upon the acoustic input sound – extends the performative arena beyond the projection screen. The work intertwines different types of semantics, so as to create a sense of causation, extend spatial presence beyond the physical space, and draw a focus upon the performer's expressiveness. There seems to be a relation between Emmerson's characterisation of the performative arena [2007], and Pressings' characterisation of sounds [1997]: Emmerson's local functions correspond to Pressings' expressive sounds, and Emmerson's field functions correspond to Pressings' environmental sounds. Pressings' notion of informational sounds can also be extended to the audio-visual domain, considering causal perception.

Environmental, expressive and informational semantics can be parameterised in any audio-visual performance language, in combination with interaction, sonic and visual dynamics, and physical setup. The parametric model of the thesis considers that the sense of spatial presence can be inferred from the combination of its parameters.

In summary, the scientific perceptual approach developed in this research seeks to benefit my practical work and be useful to any audio-visual performance practice, regardless of aesthetic approaches and technical platforms. The investigations about sonic expression and sensory dominance will lead to creative principles for the sound, the visual dynamics and the audio-visual relationship. The parametric visualisation model will articulate the variables relevant for each principle, and additional variables relevant for the audience's sense of spatial presence. I will use the model to analyse a set of audio-visual performances by different artists. By exposing how they converge and diverge from my creative principles, I will put in evidence that each principle can be explored in many different ways, and demonstrate how the model can indicate diverging creative concerns. At a practical level, the principles and the parametric model inform the development of a personal audio-visual instrument and performance language. Certain aspects of my performance work are more versatile than others, and the model will also be useful to discuss how the work's versatility complies with the creative principles.

Finally, it is important to draw the scientific boundaries and limitations of this research. The creative principles and the model are informed by scientific investigations. My practical work and analysis of other audio-visual performance languages are examples of how these investigations can benefit artistic practice and discussion. But the practical work is governed by a desire to explore rather than demonstrate the principles. The final version of my instrument explores a gray area of the principle for the image. Rather than ruling out visual dominance, the instrument behaviours lead to a discussion about how attention dynamics might change over time. This provides perspective over future research methods, beyond the aims of this thesis. As an artist interested in perception science, I am driven to confront practice and theory in ways that keep creative questions and challenges alive.

CHAPTER 3

Creative Principles

3.1 Chapter Introduction

This chapter endeavours to create operational methods for the analysis of sonic expression and sensory dominance. Accordingly, it describes three creative principles applicable in instrument design, composition and physical performance setup.

Section 3.2 elaborates on the experience of complexity, distinguishing between pragmatic and non-pragmatic modes of perception. Bridging perception science and philosophy, it stresses the importance of non-pragmatic perception for artistic experience.

Section 3.3 examines how musical instruments can convey the notion of musical expression of the thesis, introduced in the previous chapter. The section develops this notion and clarifies its boundaries, while bridging literature about music performance, interaction and NIME design.

Section 3.4 presents a taxonomy that draws from psychological and neuroscientific research, so as to relate intensity with attention, continuities and discontinuities. The taxonomy is illustrated with examples from music, audio-visual performance and animation film. I demonstrate its usefulness in the analysis of sensory dominance, outlining how the sonic dynamics can be more intense than the visual dynamics.

Section 3.5 elaborates on the relation between synchrony and perceptual binding. Drawing from perception science and audio-visual theory, I examine how different ratios of congruence/ incongruence affect perceptual prioritisation, distinguishing between conclusive and inconclusive conclusive concepts of causation.

Section 3.6 reports a study on audio-visual mapping and perceptual experience. The study draws upon methods from experimental psychology so as to inform audio-visual instrument design. My analysis/ conclusion is equally applicable to the audio-visual relationship in space.

Section 3.7 summarises the previous extrapolations in three creative principles, applicable to instrument design, composition and physical performance setup.

3.2 The Experience of Complexity: Intention vs. Intentionality

This section elaborates on the difference between pragmatic and non-pragmatic modes of perception. We have seen that events activate memories that have been previously activated by similar events [Snyder, 2001]. Only a few of these memories become highly activated and conscious; they embed unconscious memories, which Snyder calls semiactivated memories. The question is which part of the information becomes conscious, and which remains unconscious.

Gestaltist principles are a major example of how perception simplifies information according to unconscious presuppositions. Perceptual mechanics are conditioned by the rapid decay of short-term memory [e.g. Snyder 2001, Knudsen, 2007] and the primary aim of the brain [Calvert et al. 2004]. Usually perception prioritises the stimuli governed by a purpose, such as discerning a cause and a meaning, or accomplishing a task. This is typical in causal and semantic perception. Chapter 2.3.3 described these modes of perception based on Schaffer's modes of listening [1966], which Chion also considers in film [1994]. Perception science makes it clear that focusing on a goal frames the mind through previous information. Another mode of perception is possible when we are not driven by any purpose: we can also be consciously aware of a wider sensory complexity (Figure 2).

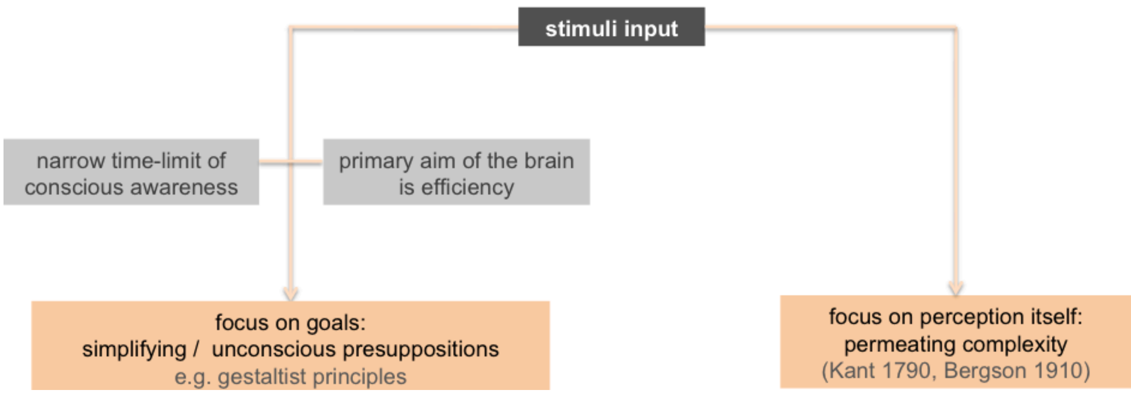


Figure 2: The experience of complexity

This broadened way of perceiving is well known in art, and there are many related philosophies. John Cage is an obvious reference, for having explored the compositional potential of chance in music [1961]. Eastern philosophies teach us that one needs to suspend pragmatic thinking in order to permeate a relation between all things. Also, Philosopher Henri Bergson saw intuition as a way to attain direct contact with a *prime reality* ordinarily masked from human knowledge [1910]. The prime reality is described as *the perpetual happening* or *duration*. It is an ongoing movement, an evolving dynamic flux that proceeds along a definite but unpredictable course. Bergson noted that the human mind is shielded from the perpetual happening by the intellect, which imposes *patterned immobility* on prime reality, separating it into discrete objects, events and processes. “In the perpetual happening itself all events, objects, and processes are unified” [Westcott 1968:8]. In Bergson's view, the intellect can freely interact with intuition to develop an enriched personal perspective. Furthermore, we can recall Immanuel Kant’s definition of *sublime* as an extraordinary experience: we fail to understand the greatness of nature by means of determinate concepts, and yet supplant this failure with a delight stemming from our ability to grasp that greatness [Kant 1790].

Conforming sensory information to concepts requires perception to segregate the information, and prioritise the converging over the diverging. In fact, every concept is an abstraction, thus a simplification. So what happens when concepts remain inconclusive? Speaking of the neural implications of artistic experience, Susan Broadhurst says: “all works of art that (…) frustrate our expectations of any clear resolution (…) are likely to activate a specific area of the frontal lobe which appears to deal with the resolution of perceptual/ experiential conflict” [2007:58]. We can say that this ‘frustrating expectations of any clear resolution’ is a unique function of art. Art can revitalise our thinking because it invites perception to bind information without subordinating to any goals. It reminds us that we can widen, narrow or dislocate our perspective, yet *every perspective is a window over reality*, as noted by Otto Roessler [1996] who is a chaos scientist specialist in hydrodynamics.

The non-pragmatic mode of perception is driven by intentionality, rather than intention. In philosophy, intentionality is described as the distinguishing property of mental phenomena of being necessarily directed upon an object, whether real or imaginary[34]. We can say that this ‘being necessarily directed upon an object’ respects to the primary aim of the brain: to make sense of the world in order to survive. But whereas intention frames the mind through previous information so as to convey conclusive concepts, intentionality places conscious awareness at ground level. The brain makes use of assumptions to simplify and clarify the perceptual field, and at the same time, it draws upon their ambivalences.

[34] intentionality, n." Oxford English Dictionary Additions Series. 1993. Online. Oxford University Press. 17 Aug. 2008

3.3 From Instrument Design to Musical Expression

The distinction between intention and intentionality is very useful to discuss how musical interface behaviours can convey a performer's idiosyncratic expression. This section looks at a set of musical instrument designs so as to examine the bilateral relation between performer and instrument, and develop the notion of musical expression introduced in chapter 2.5. Furthermore, the section presents a method to distinguish how different types of interaction convey different notions of expression.

3.3.1 The Two-Folded Interaction with Musical Instruments

Recalling our previous investigation about how different notions of flow substantiate in the interaction with musical devices, we can say that effortless interfaces convey intention, and effortful interfaces convey intentionality.

So how do interface behaviours become effortless, and how does that convey intention? For example, we use a text processor while driven by the intention of writing a text. Gaming is also all about intention – the goal of the game, the challenge of accomplishing increasingly difficult tasks, the social interaction. A person can focus on writing the text or playing the game because the interface behaves consistently. For example, if pressing the "A" key made the game player move forward in the digital 3D space, pressing "A" again should make him move forward once more. Predictability maximises our control over the interface. An interface behaving in unpredictable ways would distract us from our task; it would require attention in itself. Conversely, linear behaviours can make the interface "immaterial": the interaction becomes seemingly immediate. In the book titled *Where the Action Is: The Foundations of Embodied Interaction*, Paul Dourish described this dematerialisation of the interface as the paradigm of ubiquitous digital media [2004]; it is the paradigm nowadays as well, possibly even more [Lombard et al. 2015]. Consistent interface behaviours convey a pragmatic mode of perception, which frames the mind through previous assumptions/ concepts as described in the previous section.

Dematerialising the interface is often not desirable with musical interfaces, which can be designed to convey intentionality. Intentionality enables us to permeate a wider sensory complexity because it places conscious awareness at ground level, beyond fully predefined aims. As one dispenses with intention, interface behaviours do not need to be consistent. Indeed complex interface behaviours foster an engagement with creative expression, as shown with the study reported in "Nonlinear Dynamical Systems as Enablers of Exploratory

Engagement with Musical Instruments" by Tom Mudd et al. [2015]. The authors highlight a contrast between "communication-oriented attitudes to engagement that view the tool as a medium for transmission of ideas" and "material-oriented attitudes that focus on the specific sonic properties and behaviours of a given tool". They link material-oriented attitudes to the inclusion of nonlinear dynamical elements in musical interface design. They are particularly interested in fractal systems, where non-linear behaviours exhibit an underlying order: they are similar at different scales - the whole has the same shape as one or more of the parts.

Volatile, complex interface behaviours make the musician dwell with material properties. We can say that the interaction is bi-directional; indeed the term dwelling implies some sort of separation. This is seemingly paradoxical, because the performer also needs to embody the instrument in order to focus on the musical output: the instrument must feel like a natural extension of the body. Articulating these requirements is a compelling creative challenge in the design of new musical instruments, and it can be examined from many perspectives.

The separation of performer and instrument seems emphasised when derived from the resistances of musical systems, be they physical, or conceptual as one might find in the design of a programming language. The papers [Magnusson and Mendieta 2007] and [Bertelsen et al. 2009] show that many musicians enjoy engaging with such resistances, which exceed the performers' control. Chris Kiefer has a particular way to face the issue in live coding performances [2015]. To him, live-coding "gives the musician the freedom and power to engage very closely with digital sound synthesis processes, to create constructs and abstractions (…) less encumbered or constrained by preset and prescribed pattern, assumptions and structures". Kiefer's phrasing indicates a desire for less mediated expression. This seems difficult to achieve because code is symbolic, thus it creates a sense of material separation between the performer's agency and the musical output. To Kiefer, the separation can be excessive, particularly because computer keyboards are not designed for musical expression: "code can (…) draw the musician away from the instantaneous and engaged interactive loop that is conventionally associated with playing a musical instrument". Kiefer uses genetic programming representations to translate the output of a multi-parametric controller into code, whilst "retaining the code as the medium and the coding environment as the key focus" [Kiefer 2015]. His interface conveys both embodiment and separation: whilst the controller provides a sense of immediacy, conveying embodiment, the focus on code and the keyboard typing create a separation.

The relation between embodiment and separation is not exclusive to digital instruments. An acoustic instrument is effortful, yet the interaction becomes fluid with training, to the extent of seeming natural. At a certain point, the musician does not focus on physical gestures – their techniques became body knowledge. Unpredicted events can bring the focus back to materiality, by creating a separation between musician and instrument. For example, Pedro Rebelo speaks of a *parasitical relationship* between the grand piano and the myriad of objects used in its preparation [2015]. Introducing "parasitical" elements during performance brings unpredictability. The same can happen when audio software operates based on an audible, acoustic input: the digital sound becomes a "parasite" of the acoustic, in ways that bring unpredictability. The fact is, regardless of whether any sensors can capture the resilient nuances of the acoustic sound, software is necessarily symbolic, and physical action will always be mediated through code.

The purpose of separation is to challenge the performer's body knowledge. As the instrument "talks back", the musician is compelled to focus on its present material properties and behaviours, and the bi-directional interaction itself. The performative layer [Schroeder and Rebelo 2009] brings unique significance to the performer's actions because it is rehearsed and live. We can say that it is the combination of embodiment and material challenge that enables an "extension, a refinement, or a connection of previously untrained and unknown circuits" (Figure 3).

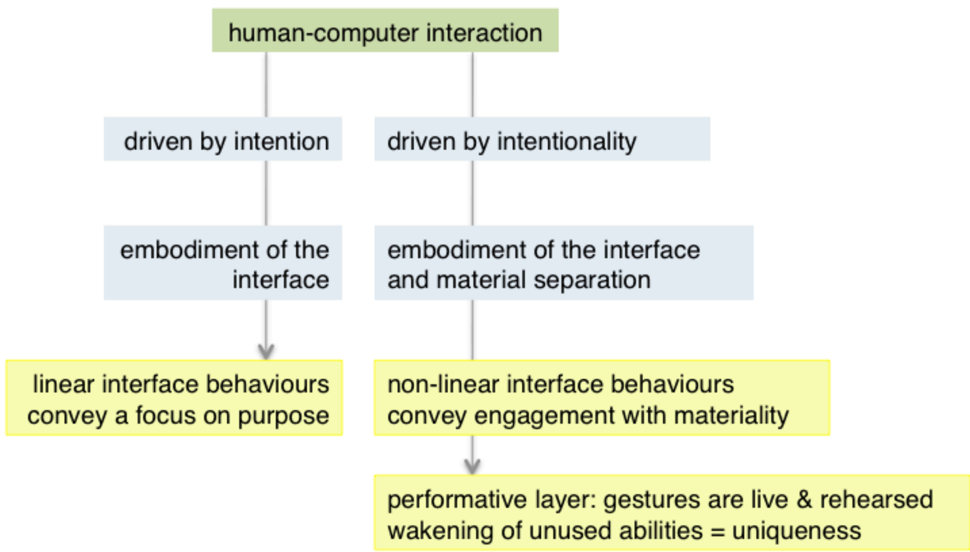


Figure 3: How linear interface behaviours convey intention and non-linear behaviours convey intentionality

3.3.2 Musical "Grid", Effort and Dynamic Deviation

A study conducted in an hospital environment showed that physical movements change from *exploratory* to *performatory* when a person becomes skilled in the execution of a specified task: movements become fluent, with a "focus on timing" [Kilbourn and Isaksson 2007]. Whilst exploratory movements imply an "initial mode of attention", with performatory movements every gesture is a "development of the one before and a preparation for the one following".

We can say that the performative layer [Schroeder and Rebelo 2009] substantiates a combination of performatory and exploratory movements, which corresponds to a combination of embodiment and separation. Whereas the performatory aspect of the music entails fluency and focus on timing, the exploratory aspect makes the musical thread unrepeatable and unique; it brings a "fresh" flavour to the music.

An effortful interface can convey these two types of movements. As a musician embodies their techniques, effort motivates a constant return to that 'initial mode of attention', a deviation from the foreseen. When asked to talk of musical interfaces and the role of effort in music, Joel Ryan observed:

> The landscape of effort runs through human bodies, our habits and our history banging up against instruments and acoustic materials. To delete effort for some idea of convenience (making it easier to make music, or for the simplicity of representation, poverty of theory) is a way to remove context from music. Effort is then a marker for the feedback between the world and our desire. [Sa et al. 2015]

Ryan pointed to this idea back in 1991, while drawing upon his work with Michel Waisvisz:

> Though the principle of effortlessness may guide good word processor design, it may have no comparable utility in the design of a musical instrument. In designing a new instrument it might be just as interesting to make control as difficult as possible. Physical effort is a characteristic of the playing of all musical instruments (…) Effort is closely related to expression (…) It is the element of energy and desire, of attraction and repulsion in the movement of music (…) Effort maps complex territories onto the simple grid of pitch and harmony. [Ryan 1991:7]

We can say that Ryan's "grid of pitch and harmony" relates to Snyder's notion of musical closure; closure entails the fulfilling of expectations derived from psychophysical processes and internalised musical traditions [Snyder 2001]. And, we can say that Ryan's "effort" reports to a dynamic deviation, which teases those expectations. It is useful to clarify how these two aspects articulate in music.

Ryan's "grid" recalls Pressing's hierarchical complexity, which refers to the existence of a structure across many levels of the music [Pressing 1983]. The "grid" becomes explicit in what Snyder described as moments of musical closure [2001]; when a musical phrase brings a sense of completion, by fulfilling expectations. These moments provide perceptual cues that lead us to structure the information in hierarchical ways, embracing diverging sensory information as a cohesive whole.

Ryan's "effort" recalls Pressing's dynamic or adaptive complexity, which refers to a rich range of behaviours over time, or an adaptation to unpredictable conditions, or a monitoring of results in relation to a reference source, or an anticipation of changes in oneself or the environment [Pressing 1983]. We can say that dynamic complexity is referential, because we are constantly comparing what we hear with the "grid" of expectations derived from our psychophysical processes and internalised musical traditions. Dynamic complexity is the constant approach and deviation from that "grid". Those deviations are often very subtle, inviting for deliberate attention. As deliberate attention increases perceptual resolution, a person becomes more susceptible to automatic attention.

In order to convey the understanding of sonic expression of this thesis, a musical instrument should assure embodiment so as to convey hierarchical complexity, and simultaneously, it should assure a level of separation so as to convey dynamic complexity.

A sonic construction reflects the performer's control over the instrument/ system, and designs can exhibit different ratios of control/ unpredictability. Chapter 2.5 showed that effort is a variable factor in musical interaction design; it can indicate if designs convey the present understanding of expression or not. I pointed out that effort should be considered relatively to cognitive demand and musical timing. Now that we looked deeper into how interaction informs musical expression we can quantify effort with respect to different understandings of expression:

- Little effort means one of two things: either the music does not depend much on the performer's interaction, or the relationship between deliberate human agency and sonic results is linear and clearly perceivable.
- Medium effort means that interface behaviours are complex, i.e. the performer needs particular skills to play the instrument; yet a sense of immediacy conveys musical timing, and/ or technical configurations rule out undesired outcomes.
- High effort implies particular skills and/ or high cognitive demand; the interaction with the system does not feel immediate, and/or the system does not rule out undesired outcomes.

The notion of expression of this thesis requires from medium to medium-high effort. It embraces a diversity of musical languages and approaches to interaction design.

3.4 Relating Intensity, Attention, Continuities and Discontinuities

Developing audio-visual work requires considering the perceived dynamics of sound and image, particularly if one is concerned with how one sensory modality affects the other. This section presents a taxonomy that can be used to analyse the sound, the image and the audio-visual relationship.

3.4.1 (Re)defining Intensity in Cross-Sensory Terms

Our concern with sensory dominance in audio-visual performance requires an operational tool for comparing the strength of sound and image. We need a notion of intensity applicable to both sensory modalities. Snyder's notion of intensity respects to music, and relates to qualities of sonic stimuli - e.g. loudness, pitch, timbre, dissonance and brightness [2001]. Perhaps we can speak of equivalent qualities in the visual domain, but each stimulus exhibits many qualities, and the intensity of events depends on how the combination is perceived, be it in the auditory or the visual domain. Snyder's wording is inspiring: intensity is any change in the stimuli chain causing an increase in neural activity [2001:62]. And his notion of musical motion is very useful: musical motion operates through an oscillation between high intensity (tension) and low intensity (release) [Snyder 2001:62].

We can adopt Snyder's definition of musical motion and redefine intensity, so as to relate intensity directly to attention. We have seen that attention can be manipulated so that audition does not subordinate to vision [Sinnett 2007]; beyond the question of sensory dominance, attention is also important in music analysis [e.g. Smalley 1986]. Attention means an increase in neural activity, and information processing should be considered relatively to the limited capacity of short-term memory, which Snyder describes very clearly [2001]. Eric Knudsen, a neuroscientist explains that the stimuli competition to access conscious awareness evaluates "signal strengths" derived from the combined effects of deliberate attention, automatic attention, and the quality of encoded information [2007].

I propose that intensity derives from this comparison: intensity is the *neural impact* of any change in the chain of stimuli causing an increase in neural activity. This notion of intensity considers the impact of stimuli, rather than the stimuli themselves. In this way, intensity depends on the event itself, on the stimuli panorama, and on the current perceptual resolution, which in turn depends on the current focus of attention (Figure 4).

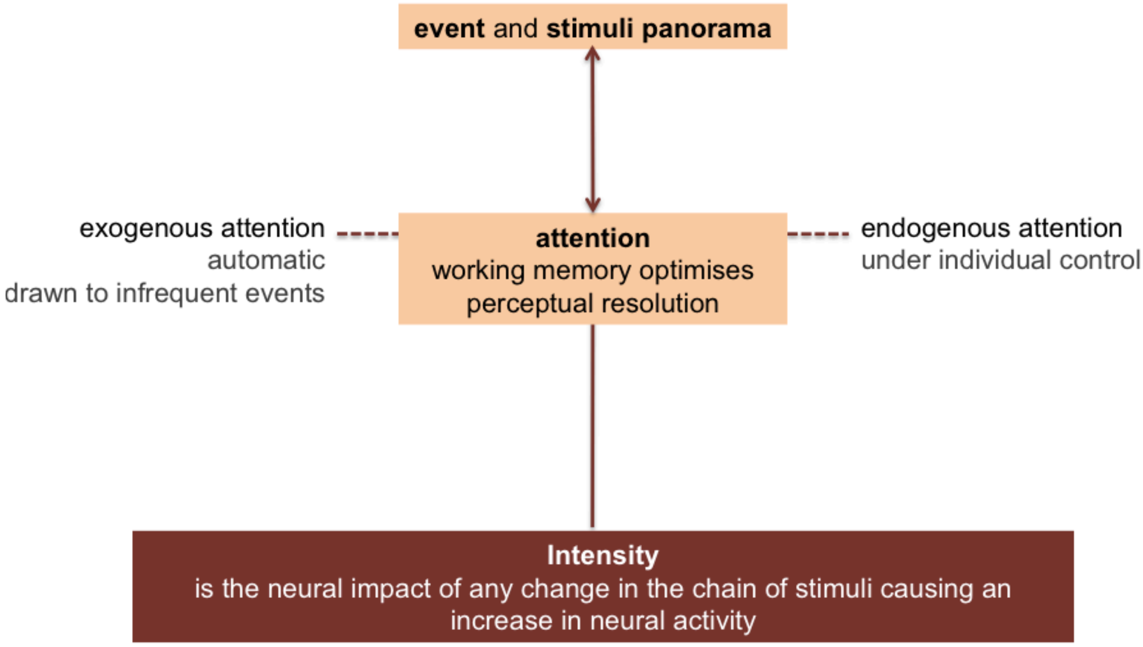


Figure 4: Intensity depends on the event itself, the stimuli panorama, and perceptual resolution.

- **Intensity depends on the event itself.**
  A change in the chain of stimuli causing an increase in neural activity is perceived as a discontinuity. Certain changes do not cause an increase unless one pays attention; in this sense, they are simultaneously continuous and discontinuous. Some types of perceived discontinuities and continuities are more variable, inter-individually, than other types. When perceiving discontinuity depends on deliberate attention, the apprehension is more variable than when a sudden event creates an interruption, prompting automatic attention. In chapter 2.4.1 we saw that the Gestaltist principles convey perceptual continuity [e.g. Wertheimer 1938, Bregman 1990, Snyder 2001]. One can perceive discontinuity, but a type of discontinuity that does not prompt automatic attention. In chapter 2.4.3 we saw that attention is automatically driven to infrequent events; the abrupt appearance or disappearance of a stimulus create great perceptual discontinuity, causing an automatic increase in neural activity [Knudsen

2007]. This strong neural impact means a high level of intensity. Furthermore, in chapter 2.4.3 we saw that automatic attention affects multisensory integration [Bertelson et al. 2000, Van der Burg et al. 2010]. This seems unavoidable, given the limited capacity of short-term memory and the primary aim of the brain: to detect, perceive and respond effectively to objects and events [Calvert et al. 2004].

- **Intensity depends on the stimuli panorama.**
  A stimulus is likely to be acknowledged (and potentially cause an increase in neural activity) when contrasting with the panorama; yet the same stimulus may be overlooked when competing with other equally bold, or bolder, stimuli. For example, a sound may stand out in a quiet environment, and disappear in a loud environment. And the colour red stands out in a black-and-white environment, but not in a yellow-red environment. Chapter 2.6 noted that the panorama also incorporates time, and that the relation between time and attention is not linear. A long period of consistent stimuli makes an inconsistent event potentially more salient than a short period of consistent stimuli. At some point, it becomes the inverse: a very long period of consistent stimuli leads us to deviate attention, so that an inconsistent event appears less salient. Furthermore, the attention target must be selected deliberately when there is considerable competition between inputs to different modalities. Chapter 2.4.3 explained that in this case, deliberate attention can produce multisensory illusions [Talsma et al. 2010].

- **Intensity depends on perceptual resolution.**
  In chapter 2.4.3 we saw that regardless of whether attention is automatic or deliberate, working memory optimises perceptual resolution for the information under focus [Knudsen 2007]. A salient, unusual event is necessarily intense as it automatically attracts attention, but a minute variation on a common event can also be intense if attention is being paid to it, and working memory optimises perceptual resolution. We saw that optimisation can happen when a person directs the eyes to a visual target, e.g. a moving image on a screen, or when the person modulates the sensitivity of neural circuits, e.g. according to music [Knudsen 2007]. The same event can be highly intense when under focus in conscious awareness, and not intense at all when ignored. We can add that deliberate attention makes a person more susceptible to automatic attention, because the attended stimuli appear more salient at high perceptual resolution.

In summary, considering the inextricable relation between intensity and attention garners a useful perspective over musical motion, be it sonic, visual, or audio-visual. Our notion of intensity considers how stimuli affect the interplay of automatic and deliberate attention. At the same time, it considers that attention influences how stimuli are perceived. We know that attention prioritises discontinuities over continuities, and that the same Gestaltist principles govern across sensory modalities. We know that sonic discontinuities are more intense than sonic continuities, and visual discontinuities are more intense than visual continuities. To compare the strength of sound and image, we solely need to consider attention dynamics relatively to multisensory integration.

3.4.2 Types of Continuities and Discontinuities

Whether musical motion is sonic or audio-visual, the interplay of continuities and discontinuities fundamentally directs attention. The intensity of each event depends greatly upon its relation with other events, as well as upon the current state of the person's attention. For example, a certain event can be more intense when preceded by a moment of low intensity, i.e. of rest. Events can also increase or decrease in intensity due to changes of perceptual resolution, rather than due to changes in the sounds and the images per se. Working memory can optimise perceptual resolution so that continuities become discontinuities, causing intensity to increase; and lowering perceptual resolution can make discontinuities become continuities, causing intensity to decrease. Systemising how different types of continuities and discontinuities affect attention is useful to the analysis of music and graphics, as well as to the analysis of sensory dominance.

I propose the following terminology, which establishes a distinction between a sense of continuity that is primarily driven from stimuli (exogenous), and a sense of continuity that depends more on the individual (endogenous). The terminology also distinguishes between discontinuities that impose disruption, and ambivalent discontinuities, whose apprehension depends on one's current perceptual resolution. Moreover, it considers how we experience cohesion in the relation between all continuities and discontinuities.

Exogenous continuity occurs when the stimuli panorama fulfils expectations; I adopt the term exogenous so as to stress that one perceives continuity regardless of individual variables such as deliberate attention and internalised traditions. Exogenous continuity can be steady or progressive:

- **Steady continuity** has no intrinsic motion; it is of lowest intensity. Being steady and continuous, it dispenses with attention. Conscious awareness is likely to deviate and focus upon any simultaneous stimuli, or upon internal states (see *Examples # 1,* bellow).

- **Progressive continuity** occurs when successive, non-abrupt events display a similar interval of motion. It fulfils the expectation that once something begins to move in a certain direction, it will continue to move in that direction - Gestalt of *good continuation* (see *Examples # 2,* bellow).

Discontinuity implies that the sequence of elements is not foreseen. Disruptive discontinuities prompt automatic attention, with an automatic increase in neural activity. Yet we can obviously also perceive discontinuity in non-disruptive ways – that is the very premise of Gestalt psychology. Hence I distinguish between radical and ambivalent discontinuity:

- **Radical discontinuity** is disruptive; it violates psychophysical expectations. It is of highest intensity, prompting automatic attention (see *Examples # 3,* bellow). Radical discontinuities "monopolise" short-term memory. They are prioritised in any stimuli competition to reach conscious awareness, and the increase in neural activity means great demand in terms of information processing.

- **Ambivalent discontinuity** is simultaneously continuous and discontinuous. Perceiving discontinuity depends on deliberate attention – on working memory optimising perceptual resolution. At lower resolution, the foreseeable logic is shifted without disruption. At high resolution, the discontinuity becomes more intense. Higher intensity implies greater attention/ neural activity, and lower intensity implies lesser attention/ neural activity (see *Examples # 4,* bellow).

Finally, **endogenous continuity** "wraps" perceptual motion, binding any types of continuities and discontinuities in meaningful ways. It occurs at a high hierarchical level in perceptual organisation. This is perhaps the most subjective type of continuity. It depends on complex synergies of biological/ psychophysical factors, combined with personal and cultural factors. I use the term endogenous so as to stress that perceiving cohesiveness depends greatly on the individual. Endogenous continuity corresponds to the mental representation of the work as whole.

We can say that every artistic work aims the audience to experience endogenous continuity. But all other described continuities and discontinuities can be illustrated with paradigmatic examples:

Examples # 1: steady continuity

A musical example of steady continuity is Elaine Radigue's *Triologie de la Mort* [35], a three-hour musical drone work with little amount of overtones/ harmonic variation. The work follows the path of the continuum of the six states of consciousness according to Tibetan Buddhism. An audio-visual example of steady continuity is La Monte Young and Marian Zazeela's *Dream House* (begun in 1962) [36], a long-lasting and architecturally adaptive installation [see Centre Pompidou 2005]. The work defines a vibratory space through the combination of continuous sound frequencies and continuous light frequencies, experimenting on how people are drawn to inhabited it. Rather than providing an object, this work creates circumstances for contemplative, undetermined experience.

Examples # 2: progressive continuity

In the sonic, visual and audio-visual domain, every gradual increase or decrease in intensity creates progressive continuity – e.g. in loudness, tonality (e.g. musical scales and harmonic progressions), brightness, colour, density or duration. For example, Gary Hill's *Black and White Text* [37] creates progressive continuities in rhythm and density: the intervals between the words become progressively shorter, as sound layers accumulate. Importantly, we cannot speak of progressive continuity unless we perceive motion. If the progression happens so slowly that we cannot immediately apprehend any change (as happens in Radigue's *Triologie* for example), we should rather speak of steady continuity.

Examples # 3: radical discontinuity

Radical discontinuities can be used so as to create rhythmic patterns, as happens in Vasulka's *Heraldic View* [38], or in club music. The sequence of elements fulfils expectations, as happens with any pattern; but although we perceive continuity at a structural level, each abrupt event will inevitably cause an increase in neural activity. Alternatively, radical discontinuities can be explored so as to tease and counterpoint expectations, be they of psychobiological nature, and/or derived from internalised traditions. In *Corona*, a CD by Pan Sonic[39] (originally called Panasonic), the duo Mika Vainio & Ilpo Väisänen uses several radical discontinuities. And as an audio-visual example, we can refer to the sudden blackouts from Ryoji Ikeda's performance *Superimposition*[40].

---

[35] CD released in 1998 by Niblock's Experimental Intermedia Label
[36] https://www.youtube.com/watch?v=3ahgq-zVQLc
[37] https://www.youtube.com/watch?v=bg1O3NcPwBg
[38] http://www.fondation-langlois.org/html/e/page.php?NumPage=481
[39] CD released by Blast First Petite, 2010: https://www.youtube.com/watch?v=kPhkH1da08o
[40] https://vimeo.com/49873167

Examples # 4: ambivalent discontinuity

Thomas Wilfred's *Lumia*[41], performed with the *Clavilux,* stressed polymorphous, fluid streams of colour slowly metamorphosing, inviting attention to focus on subtle detail changes. And typically, minimalist music / art works aim us to discern, and focus upon subtle variations in that which at first sight appears continuous. For example, in Phill Niblock's film/ performances series *Movements of People Working*[42]*,* the repetitive movements of manual labour invite us to abstract from the labour, and from who's performing the labour. The continuous sonic mass of sound invites perception to focus on the sonic and visual textures; as we modulate the sensitivity of neural circuits, the microtonal, harmonic variations become intense.

3.4.3 Using Continuities and Discontinuities

In many musical languages, sounds relate in surreptitious manners. Irregular and multi-layered, the sonic motion ties and unties sonic qualities at many hierarchical levels, inviting perception to navigate between detail and structure. That happens often in free jazz, for example, or when many musicians play freely together, as in the Variable Geometry Orchestra [43]. The sonic orchestration intertwines the whole range of continuities and discontinuities, fulfilling, tempting or violating expectations. Whilst that which Pressing called dynamic complexity [1986] keeps deviating our expectations, deliberate attention enables us to perceive and enjoy an underlying structure, i.e. Pressing's hierarchical complexity [1986]. It is useful to recall Snyder's description of partial closure (chapter 2.4.1): partial closure creates expectations of eventual closure [Snyder 2001]. A subsequent, more completely closed musical phrase will also close any preceding, less closed phrases. We can experience endogenous continuity because we make sense of how the music articulates dynamic and hierarchical complexity. But that requires attending how the aural unfolds: how a moment in the musical motion reports to a multitude of previous sounds, silences, or kaleidoscopic soundscapes.

If we close our eyes and modulate the sensitivity of neural circuits so as to perceive the surreptitious relations between the sounds, perceptual resolution increases and decreases according to the sonic motion. If the desire is to maintain a similar perceptual oscillation while the audience is directing the eyes to a moving image on a screen, the work must dispense with radical visual discontinuities. These would automatically attract attention, influencing multisensory integration and subordinating audition. We have seen in chapter

[41] https://www.youtube.com/watch?v=ojVX8FWYc4g
[42] DVD published by Microcinema, 2003
[43] https://www.youtube.com/watch?v=yZT42pMph2g

2.4.3 that the image pops out when radical visual discontinuities are synchronised with radical sonic discontinuities [Van der Burg et al. 2010]. And we have seen that non-synchronised, radical visual discontinuities disrupt the sense of auditory continuity [Kobayashi 2007].

I do not mean that audio-visual performance works exhibiting radical visual discontinuities are less concerned with musical experience. For example, Rioji Ikeda is primarily a musician, and yet he explores radical discontinuities synchronised one-to-one. His work benefits from this because it is all about manipulating perceptual experience; it never leaves attention under individual control. The audio-visual motion explores a high level of intensity.

In other works, the sonic and visual motions have a similar amount of tension and release i.e. variability in speed and form, but attention is left more under individual control. This is the case in John Whitney's work [1980]; or in *The Magic Sun*, Phill Niblock's film[44] with Sun Ra, Marshall Allen, John Gilmore and Danny Davis; or in *7-of-12 dialectologies*[45]*,* an audio-visual performance by Daniel Schorno and Haraldur Karlsson.

Regardless of whether the image exhibits radical discontinuities or not, a same level of tension and release in sound and image can be problematic when the music unfolds in complex, multilayered and often rapid ways; and perceiving the logics of sonic motion requires undivided attention. If one desires attention to focus upon the relations between the sounds themselves, one can opt for exploring a limited set of visual variations within steady visual continuity. An example is Thor Magnusson's *Threnoscope*[46] system, which is made for drone music; the projected image consists of a graphical representation of the sounds, in a two dimensional, circular score [Magnusson 2014].

But steady visual continuity is not mandatory. The diagram in Figure 5 shows another way of articulating continuities and discontinuities. The sonic motion oscillates between points of low and high intensity. In-between steady continuities and radical discontinuities, there are gradients of ambivalence, where experiencing continuity or discontinuity depends on deliberate attention. Attention can focus on the wealth of sounds if Gestaltist principles convey the perceptual simplification of visual dynamics. This is the case with progressive continuities and ambivalent discontinuities, which do not prompt automatic attention.

---

[44] http://www.dailymotion.com/video/xmmkj0_sun-ra-magic-sun_shortfilms
[45] https://vimeo.com/27694220
[46] Video footage from performance at the ICLI 2014 conference: https://drive.google.com/file/d/0B9f9deWrFgQuSU1NbFJmdkZWMU0/view?usp=sharing & rendering of the *Threnoscope* system: https://vimeo.com/75380587

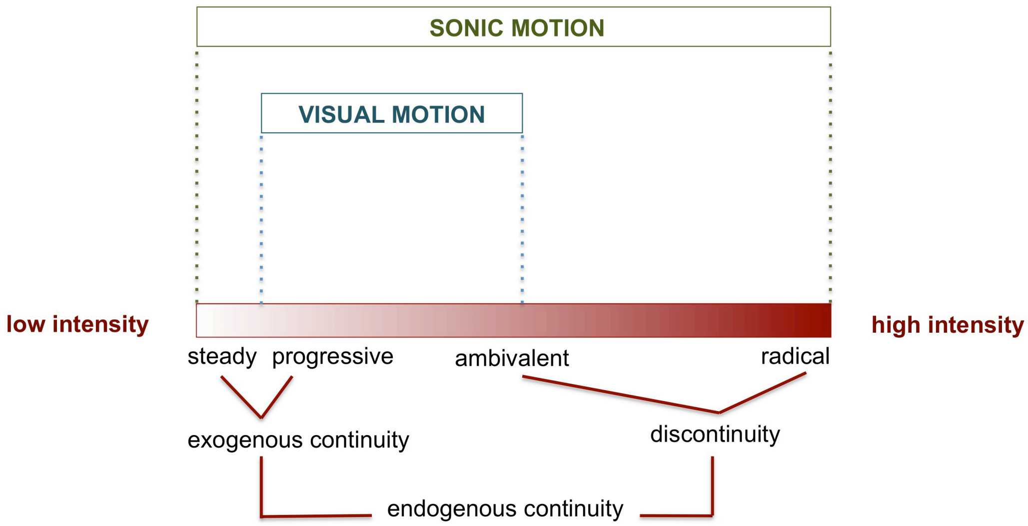


Figure 5: Articulating the sonic and the visual motion such that the experience is driven through sound organisation

If we want to convey the perceptual simplification of visual dynamics, we need to consider the moving image as a whole. While multiple visual changes may occur at once, they should converge into a same form. In VJing, the moving image often orchestrates a number of different shapes – musical voices – that form polyphonic movements; each voice creating as much, frequently more tension and release than does sound. This is demanding for the limited capacity of short-term memory; we can remember 3-5 chunks of information at the time [Snyder 2001, Cowan 2001, Gilakjani 2012]. There is a great amount of neural activity involved in perceiving each separate 'voice', and making sense of the visual polyphony. The demand for visual information processing could obfuscate the sonic experience, blinding perception to the nuances of sound organisation.

Alternatively, one can control the amount of mental activity that the visual information imposes on working memory. A monophonic moving image can bring an environmental context to the sonic experience, punctuating the sonic motion in non-disruptive ways (Table 2*).* At times visual details can become intense; we have seen that this is possible with deliberate attention. However, that depends on the current perceptual resolution, which can be driven by the relations between the sounds themselves.

Sound	Moving image	Relative intensities	Results
Steady continuity	Progressive continuity Ambivalent discontinuity	The aural is less intense than the visual	Because neither sound nor image are disruptive, perception is invited to draw upon ambivalent discontinuities
Progressive continuity Ambivalent discontinuity	Progressive continuity Ambivalent discontinuity	The aural is as intense as the visual	Because neither sound nor image are disruptive, perception is invited to draw upon ambivalent discontinuities
Radical discontinuity	Progressive continuity Ambivalent discontinuity	The aural is more intense than the visual	Because the sound is disruptive and image is not, attention is automatically driven to sound

Table 2: Perceptual integration of the sonic and the visual motion

3.5 Synchrony, Perceptual Binding and Fungible Audio-Visual Relationships

The perceptual simplification of visual dynamics is advantageous for complex sonic constructions, but simplifying the audio-visual relationship might not be desirable. Many composers [e.g. Schaeffer 1966, Pressing 1997:10] noticed that perception tends to bias those sonic events and qualities which help visual apprehension. It tends to subordinate other qualities such as timbre, texture, vibration, and the nuances of the performer's expression, which form the wealth of multilayered relations between the sounds themselves. Chapter 2.4.4 highlighted that in perceptual binding the visual discounts the aural and the aural discounts the visual [Kubovy and Schutz 2010]. The discount is of course conditioned by the primary aim of the brain [Calvert et al. 2004]; humans need to detect, perceive and respond to causation in order to survive. As audition supports vision, the audio-visual tends to skew visually [e.g. Colavita 1974, Pressing 1997:10]. But visual dominance is not mandatory. Sinnett et al. observed that even though the visual channel is sampled before, or more often than the auditory channel, the difference in sample rate is attentional in nature, hence attention can be manipulated so that vision does not dominate audition [2007]. He did not say how. The taxonomy described in the previous section is useful to assess on how discontinuity/ continuity affect the interplay of automatic/ deliberate attention, but it does not fully resolve the problem. The question now is, must perceiving causation really be prejudicial to the experience of music in audio-visual performance? Indeed, an audio-visual connection brings additional meaning to the audience's experience.

The relation between audio-visual synchrony and perceptual binding is multifaceted. For a start, the term synchrony requires clarification. For example, Stevens describes synchrony as a regular process of recurring accents of music and image that appear to occur at the same time [2009]. In music accents would be made with tone, loudness or beat; in image, with movement or change. Stevens' use of the term becomes confusing because sometimes it refers to perceptual binding, and other times to mechanical connection. Her *neutral synchronisation* occurs when "there is no synch but nothing is out of synch". Her a*pparent synchronisation* occurs when "sound creates apparent visual movement, and there is an expectation of movement but none actually occurs". And also, when "synchronisation occurs within a window of time where it appears synchronised but when measured exactly it is not". This thesis confines the term synchrony to machine-synchronisation, distinct from perceptual binding.

In film theory, Chion states that synchrony leads inevitably to perceptual binding [1994]. In experimental psychology, Kubovy and Schutz state that it does not [2010:57]. This seems in contradiction, but perhaps it is not. Film invites us to suspend disbelief, and perception can be driven to bind information in unusual ways. For example, the sound of breaking glass increases the dramatic strength of a character dying in the arms of another. In everyday life our mode of perception tends to be more pragmatic.

Recalling the study from [Schutz and Kubovy 2009], when sound and image were desynchronised the perceived marimba sound was shortened so as to coincide with the visible impact, but the piano sound was perceived in full length. The perceived sound was shortened when the visible impact preceded the sound, and not when the stimuli sequence was reversed, nor when the stimuli were synchronised. When we bind a marimba sound with the visible impact that seems to have caused it, the perceptual representation conforms to previous concepts of causation. Asynchrony is incongruent with those concepts, but the audio-visual relationship is otherwise so plausible that one overlooks that incongruence so as to produce an integrated audio-visual representation.

Chion wrote that synchrony is more determining in audio-visual binding than verisimilitude: "To the spectator, it is not the acoustic realism so much as synchrony above all, and secondarily the factor of verisimilitude (verisimilitude arising not from truth but from convention), that will lead him or her to connect a sound with an event or detail" [1994:23].

Also, the term synchresis refers to "the forging of an immediate and necessary relationship" between synchronised sounds and images, whether their combination is plausible or implausible [Chion 1994]. We can say that a person binds sound and image because it is implicit that they have a common origin: the film. As film frames attention, synchrony invites us to discern an audio-visual linkage, even if the link seems incongruent.

Boltz observed that congruent sound/image pairings lead to integrated perceptual encodings and memory representations, and incongruent pairings lead to separated encodings and representations for each sensory modality [2004]. This conclusion is consistent with Kubovy and Schutz' conclusions in experimental psychology [2010], and it is very useful for examining the role of synchrony in film and audio-visual performance. Whether the meanings of the sounds are congruent with the meanings of the images or not, synchrony creates a level of audio-visual fit, i.e. congruence, which can be used to create incongruent, that is implausible relationships. Binding sounds and images while finding the combination implausible is creating new memory chunks based on separated perceptual encodings and memory representations. The resulting perceptual representation is possibly more instable than when perception skews sensory information so as to forge an integrated audiovisual representation that conforms to previous memories.

The strength of perceptual binding depends on our assessments about ecological fit, i.e. congruence or plausibility, and those assessments are based on concepts of causation. We can relate Kubovy's and Schutz' notion of an audiovisual object [2010] with Snyder's notion of unconscious, semiactivated memories [2001]. Humans bind sound and image when the combination activates unconscious memories in which the auditory and the visual stimuli had a same physical origin, i.e. a same cause. Multisensory integration and Gestaltist processes are driven by this very assumption - the common cause of stimuli. In both cases, the meaning of the parts derives from the whole. The common thread from low level to high level in perceptual organisation is to detect, perceive and respond effectively to the world, and perceiving the cause of stimuli in good time is crucial to survive.

It is useful to recall that stimuli can be detected without ever reaching conscious awareness. Spence proposes a good explanation for the experiments about the Colavita visual dominance effect, where the participants have to rapidly discriminate between auditory and visual stimuli [2009]. He argues that the participants might detect the auditory stimuli, and initiate their response "prior to becoming aware of the stimuli eliciting that response" [2009:254]. Their awareness of stimuli would then be modulated by their response. Vision-only responses would prevail, not because of an attenuation of auditory input, but because the visual channel is usually sampled more often than the auditory channel (Spence supports this with [Sinnett et al. 2007], which shows that the sampling rate depends on attention).

We should consider the perceptual process in an integrated manner when examining how ecological fit affects perceptual binding and attention. I do not aim to distinguish if sensory interactions occur at lower or higher level in information processing. Since the question is how perceived causation affects the experience of music in audio-visual performance, we solely need to consider ecological fit in terms of perceptual prioritisation and cognitive workload.

Perception prioritises information according to interpretative frameworks. For example in Kubovy's experiment, perception shortens the marimba sound because the resulting audio-visual representation fits with a clear concept of causation; the ecological fit is conclusive. The Gestaltist principles also describe how attention is driven to prioritise sensory information that fits with concepts; non-fitting information can be perceived and ignored, in favour of a congruent representation. The clearer is the interpretative framework, the lesser cognitive workload is required to process sensory information: congruent stimuli require less cognitive processing than incongruent stimuli [Brown and Boltz 2002]. This is because congruence conveys perceptual chunking. As congruent stimuli conform to concepts, perception chunks the information so that short term-memory handles the integrated representation through cues.

This understanding justifies a particular definition of high ecological fit, which can be used to analyse audio-visual performances: we can speak of high fit whenever the audio-visual relationship produces conclusive concepts of causation. Conclusiveness can happen even when inconsistent information reaches conscious awareness, as shown in Gestaltist psychology. It leads perception to chunk the information according to previous concepts, and neural circuits are modulated accordingly. High fit is not demanding in terms of cognitive processing; it implies interpretative continuity. Continuity dispenses with attention, and we may solely take in consideration those details of an experience that fit our prejudicial image of what the event should be. Hence, a high audio-visual fit can be problematic for the experience of complex musical constructions.

Attention works differently when the stimuli do not fit any concept. About incongruent sound/ image pairings Boltz observed: “it is not always clear where attending should be directed or how the conflict of information can be resolved within one interpretative framework" [2004:1196]. Her research shows that a high level of incongruence leads perception to forge separate auditory and visual representations. We can say that perceptual binding is weak because the sound/ image pairing does not activate previous memories of causation. That can be a measure for low ecological fit. Nevertheless, in audio-visual performance and film we are encouraged to discern relations between sound and image. Highly incongruent relationships create interpretative discontinuities that are demanding in terms of cognitive processing; they cause a great increase in neural activity. Given the limited capacity of short-term memory, attention is likely to focus on making sense of the audio-visual relationship, rather than on making sense of the relations between the sounds. That can be problematic for the music, inclusively because attention plays a central role in temporal experience [Brown and Boltz 2002].

Between congruence and incongruence, there must be a medium level of audio-visual fit, corresponding to ambivalent concepts. Gestaltist psychologists have described this in the aural and the visual domains; the term *multistability* refers to when perception pops back and forth between different interpretations [Rubin 1921]. Extending Kubovy's and Schutz' notion of an audiovisual object, we can speak of inconclusive audiovisual objects to describe perceptual representations that are not really congruent or incongruent. The audio-visual relationship appears congruent enough to convey perceptual binding, and incongruent enough to convey separate representations for each sensory modality. In any case, attending one sensory modality should affect how we perceive the other modality: attention determines the perceptual scale of each representation, and consequently, the ratio of perceived continuities/ discontinuities.

If we desire the image to create a reactive stage scene, there must be a sense of causation, which requires one to form audiovisual objects. And if we desire the audience's attention to focus on the relations between the sounds, those objects must be inconclusive. The challenge is to create an audio-visual relationship that produces causal percepts, but also throttles the fit between the sonic and the visual events, so that these percepts remain inconclusive. I call this a **fungible audio-visual relationship**. The fungible relationship exhibits medium ecological fit, conveying inconclusive audiovisual objects.

With a fungible audio-visual relationship, some semiactivated memories (using Snyder's term [2001]) drive to us to detect causation, conveying perceptual binding. Yet others are in contradiction; binding would be very weak if solely those had been activated. This happens at unconscious level; since conscious awareness does not segregate the fitting information from the non-fitting, concepts of causation remain inconclusive. Each sonic and visual event is a different perceptible event, so they are each individually assessed when in working memory. They are only chunked when not under attention, and the cause and effect relationships are not checked for. Points of sensory unison create an audio-visual connection, but it becomes unclear whether the connection is purely perceptual, or if it is based on mechanical cause and effect relationships. As the audience senses causation, they may seek to understand the instrument. As the cause-effect relationships become confusing, they are invited to explore perceptual experience itself.

3.6 Study About Audio-Visual Mapping and Perceptual Experience

This section reports a study that demonstrates how an audio-visual mapping can create a fungible audio-visual relationship. The study demonstrates the fungible mapping regardless of aesthetic approaches and technical platforms. It has two complementary purposes: to inform instrument design, and to expose differences between my definitions of low, medium and high ecological fit.

The study draws upon methods from experimental psychology. Baptiste Caramiaux (research assistant, Computing, Goldsmiths, London University) introduced me to the procedures of controlled experimentation, and made statistics tests of quantitative data (T-Test). I designed the experiment, created audio-visual mapping prototypes, and analysed quantitative and qualitative data. The results will be discussed considering Gestaltist principles and the notion of audiovisual object [Kubovy and Schutz 2010], which the previous section extended with a distinction between conclusive and inconclusive objects. In this way, conclusions will also be applicable to the audio-visual relationship in space.

3.6.1 Problem

Previous researches showed that low ecological fit conveys weak perceptual binding, and high fit conveys strong binding [Boltz 2004, Kubovy and Schutz 2010]. The notion of fungibility requires us to demonstrate what happens with a medium fit. I proposed that we can speak of high fit whenever concepts of causation are conclusive, whether inconsistent sensory information reaches conscious awareness or not. The difference between high and medium fit can be validated if we demonstrate the difference between conclusive and inconclusive concepts of causation. So how can we demonstrate a fungible audio-visual mapping, and how is it perceived?

3.6.2 Hypothesis

A fungible mapping should be congruent enough to create a sense of causation, and incongruent enough to confound the cause and effect relationships; thus it might combine mappings that convey a sense of causation, and mappings that do not. We wanted to see how complexity affects the clarity of perceived cause and effect relationships.

Do cause-effect relationships remain clear if they are inconsistently interrupted?
We could start with the assumption that synchrony produces a sense of causation. It conveys a Gestaltist principle called common fate, which manifests when we group simultaneously changing elements. We expect that when an object moves, all its parts move together; thus we bind sound and image if they change at the same time. In addition to common fate, consistent synchrony conveys Gestalts of good continuation: it fulfils the expectation that when something begins to move in a certain direction, it will continue to move in that direction. Aslo, people are generally familiar with audio-visual software and VJ culture, which means that consistent synchrony produces an impression of clearly perceivable cause-effect relationships. We wanted to see if these relationships remained clear once synchrony was inconsistently interrupted; the interruptions should be clearly noticeable. Clearly perceivable cause and effect relationships would mean conclusive concepts of causation, and high ecological fit.

Does random latency produce a cause and effect relationship?
We decided to see how latency, i.e. the delay between the audio and visual stimulus, affects perceived causation; a weak sense of causation would mean low ecological fit. And what if we add more perturbation? Using Schutz' and Kubovy's marimba experiment as a guide [2009], we compounded perturbations by 1.) randomising latency, 2.) randomly interrupting the cause-effect relationship, and 3.) adding the perturbation of a synchronised, not interpolated visual parameter.

Given that synchrony conveys a sense of causation, does the feeling persist when complexity obfuscates the base cause-effect relationships?
That should manifest as a quantifiable gap between the participants' sense of causation and their sense of understanding the cause and effect relationships. Such a gap would be characteristic of medium ecological fit.

To grasp underlying perceptual dynamics, we asked for a description of the perceived cause and effect relationships. We could analyse the qualitative answers by considering Gestaltist principles [Wertheimer 1938, Bregman 1990, Snyder 2001] and the notion of audiovisual object as mind-dependent concept [Kubovy and Schutz 2010].

3.6.3 Stimuli

The study employed four audio-visual mapping prototypes, programmed in Processing, the Java-based procedural graphics environment[47], and shown on a computer. The same audio recording was used in all prototypes: a short orchestration of string instruments (37 seconds), with amplitude ranging between 0 and 43 (arbitrary values, in linear scale).

We dispensed with computer-generated sounds, whose timbre would fit with computer-generated images. In this way, we ensured that binding was due to the mapping, independently from the qualities of the sounds and the images. We did not need any mapping prototype to exhibit maximum fit in order to distinguish between high and medium fit, because that distinction should derive from the difference between conclusive and inconclusive interpretations of causation.

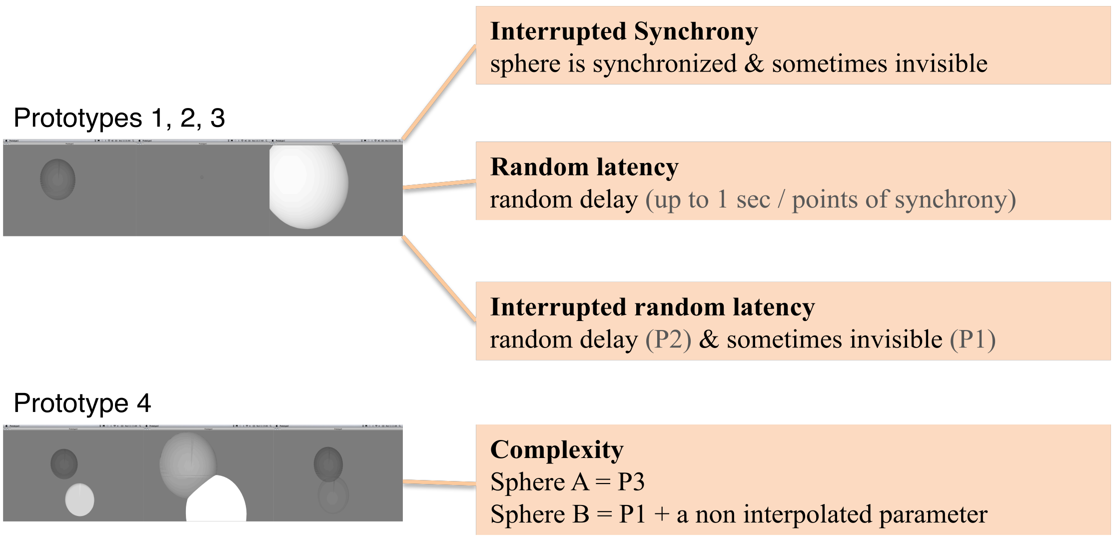


Figure 6: Prototypes 1, 2 and 3, which exhibit one sphere; and Prototype 4, which exhibits two spheres (http://doc.gold.ac.uk/~map01apv/StudyMappingPrototypes.mp4)

[47] http://processing.org/

The prototypes were black and white. Prototypes 1, 2 and 3 exhibited a single sphere drawn in a digital 3D space; prototype 2 exhibited two spheres (see Figure 6). In all prototypes, audio amplitude modulated the spheres' size, colour/ transparency, and position. In Prototypes 1, 2 and 3, all visual parameters were slightly interpolated, which smoothened otherwise frantic visual changes. In Prototype 4, the sphere's position was not interpolated, which made its visual aspect more frantic. Indeed, the detected amplitude changes are faster and bigger than how we perceive the audio-visual composite. In this sense, the result of interpolation corresponds to perceptual simplification: it enables clearer visual apprehensions.

Prototype 1 / Interrupted synchrony

In Prototype 1 all parameters from the sphere are synchronised with amplitude detection of the audio input. The sphere is visible between amplitude 0 and 6, and amplitude 19 and 43; it is invisible between amplitude 7 and 18 (this interval is in the range of average amplitude values).

Prototype 2 / Random latency

In Prototype 2 the sphere is drawn with random delay upon amplitude detection. There are occasional points of synchrony, and maximum delay is 1 sec (multisensory interactions due to plausible cause-effect relationships may occur even when the effect is delayed up to 700msec [Kubovy and Schutz 2010]).

Prototype 3 / Interrupted random latency

In Prototype 3 the sphere is drawn with random delay upon amplitude detection, as in Prototype 2. In addition, it is invisible between amplitude 7 and 18 (i.e. it is visible solely when amplitude is between 0 and 6 or 19 and 43).

Prototype 4 / Complexity

Prototype 4 displays two spheres that merge and split. Sphere A is drawn with random delay upon amplitude detection, as in Prototype 3. Sphere B is synchronised and invisible between amplitude 7 and 18, as in Prototype 1. Because the position parameter is not interpolated in sphere B, the sphere moves frantically through the X and the Z-axis in the digital 3D space.

3.6.4 Procedure

The ten participants were recruited from Goldsmiths, University of London. All had knowledge of computing. Thus, if they did not understand a mapping (low Transparency rate) and yet felt causation (high Causation rate), it could not be due to being unfamiliar with software. Importantly, at the time nobody was acquainted with my investigations about cognition/ attention; they were published later.

The experiment included ten individual sessions. Firstly, the participants were asked to read a questionnaire; they should respond after viewing each of our four audio-visual mapping prototypes. Then they were played the audio recording alone, prior to viewing the prototypes in random order. The questionnaire asked them to rate their sense of causation (Causation) and their sense of understanding the cause-effect relationships (Transparency), on a Likert scale (between 1 and 7). Additionally, it asked them to explain their Transparency rates.

The questions were posed as follows (the headers added here, prior to the questions, were not included):

A. Causation / quantitative question:
How would you scale the sense of a cause-effect relation between sound and image? (1 = no cause-effect relation between sound and image; 7 = all visual changes relate to changes in sound)

B. Transparency / quantitative question:
Can you distinguish which input factors affect which output parameters, and how? (1 = never; 7 = always)

C. Explanation / qualitative question
Please explain your rating in question B.

The average ratings for each mapping prototype were calculated, generated statistics, and compared. Our analysis and discussion make use of the answers to the qualitative question (C).

3.6.5 Results

Table 3 shows the average over the participants' ratings for Causation (question A) and Transparency (question B).

PROTOTYPE	A/ causation	B/ transparency
1	5.6	5.1
2	2.7	2.3
3	1.8	1.9
4	4.7	2.8

Table 3: Average ratings for Causation and Transparency

We used a statistics test (T-test) to compare the ratings. It showed that:

- The subjective rating for Causation is significantly higher in Prototypes 1 and 4 ($p < 0.05$) than it is in Prototypes 2 and 3 ($p < 0.05$).
- The subjective rating for Transparency is significantly higher in Prototype 1 than it is in Prototypes 2 ($p < 0.05$), 3 ($p < 0.05$) and 4 ($p < 0.05$).
- The gap between the subjective ratings for Causation and Transparency is not significant in Prototypes 1, 2 and 3 ($p > 0.05$).
- The gap between the subjective ratings for Causation and Transparency is significant in Prototype 4 ($p < 0.05$).

The statistic test establishes a distinction between “high” or “low” rates. Figure 7 represents these results.

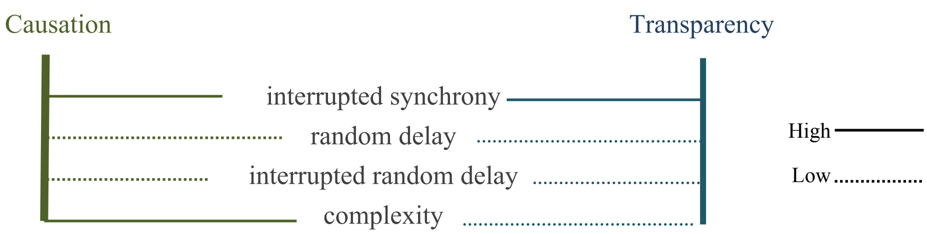


Figure 7: Relation between Causation and Transparency for each mapping prototype

The aspect of the prototypes can be inferred from the answers to question C, which aimed the participants to explain their Transparency rates.

Both Causation and Transparency were rated high in Prototype 1/ Interrupted Synchrony. Six participants wrote of one changing visual shape, in the singular. For example, a person who rated Causation 7 and Transparency 7 wrote: “sound affects sphere rhythm and size”

[Participant #7]. The other four participants spoke of two visual shapes. For example, a participant who rated Causation 7 and Transparency 6 wrote: “the higher the frequency, the bigger the circle; each touch of zither makes a small circle” [Participant #10].

In Prototype 2/ Random Delay, the global rating was low for both Causation and Transparency. Four participants wrote that they did not find any audio-visual relationship. Other participants did. For example, a person who rated Causation 2 and Transparency 2 wrote: “I have the impression that amplitude affects the shape, but not all consistently” [Participant #8]. Another person who rated Causation 3 and Transparency 2 wrote: “amplitude affects nothing (or little); it seems the guitar is affecting, but it proved inconsistent; some delay effect” [Participant #9].

Similarly, the global ratings were low in Prototype 3/ Interrupted Random Delay. Six participants did not find any audio-visual relationship. For example, a person who rated Causation 1 and Transparency 1 wrote “no relation” [Participant #4]. The other four persons did find a relationship. For example, a person who rated C2 and T2 wrote: “at times when the sphere is small it synchs to sound, at other times it appears during pauses between musical phrases” [Participant #1].

In Prototype 4/ Complexity, causation was rated high and transparency was rated low. Several participants stated that they found a relation between sonic and visual events, and that this relation was confusing. For example, a participant who rated Causation 7 and Transparency 1 wrote “no idea” when asked to explain the transparency rate [Participant #5]. Three participants did not distinguish any specific cause-effect relationship, and the others assigned multiple visual parameters to multiple sonic parameters. Among these, four specified that they were uncertain. For example, a person who rated C6 and T3 wrote: “it is hard to concrete one input with one output. There are many possible factors, but some seemed to have gotten the sound very closely” [Participant #3].

In this prototype Complexity, one component mapping equals the one from Interrupted Random Latency (sphere A, in Complexity). Thus we compared the data from these two prototypes. After viewing Interrupted Random Latency, a participant who rated both Causation and Transparency 1 explained: “sound and image do not synch”. Yet after viewing Complexity, the same person rated Causation 7 and Transparency 5, explaining: “rhythm and tempo affect the sphere; all visual changes are produced by sound” [Participant #7]. This means that the person saw a relation between the sounds and the visual shape from Interrupted Synchrony; yet when viewed independently, this shape had been assessed to exhibit no relationship with sound.

That person saw Interrupted Random Latency prior to Complexity, and one could ask if the change was due to the order in which the prototypes were shown. Yet, we found a similar change with another participant, who saw Interrupted Random Latency after Complexity [Participant #4]. In Interrupted Random Latency, the person rated Causation 2 and Transparency 1: “I can't identify anything”. In Complexity, he rated Causation 4 and Transparency 3: “the instrument type effects position; amplitude effects size; some delay, too?” Whereas in Interrupted Random Latency he detected no cause-effect relation, in Complexity he detected mechanical delay.

3.6.6 Analysis and Discussion

Prototype 1/ Interrupted Synchrony aimed at confirming that audio-visual synchrony conveys causation, and at seeing whether the cause and effect relationships remained clear despite being inconsistently interrupted. Prototype 2/ Random Latency tested whether random delay produces a sense of audio-visual relationship. Prototype 3/ Interrupted Random Delay was meant to test the same; except it compounded inconsistencies, by randomly interrupting the cause-effect relationship. Finally, Prototype 4/ Complexity aimed at testing whether the sense of causation persists if one combines synchronised and non-synchronised audio-visual components, with a multitude of perturbations. Complexity combined the mappings of Interrupted Synchrony and Interrupted Random Delay, plus an additional mapping also exhibiting synchrony, but with lesser interpolation.

The participants found Causation high in two of the Prototypes: Interrupted Synchrony and Complexity. This indicates that synchrony is taken to reveal a cause and effect relationship. Conversely, the rating for Causation was low in both Random Delay and Interrupted Random Delay. In both prototypes, several participants perceived some cause and effect relationship, possibly due to the existing points of synchrony. Yet the low rating for Causation shows that perceptual binding was too weak to be convincing. In other words, *the sense of causation is weak when points of synchrony are sparse*.

In Interrupted Synchrony, the cause and effect was interrupted for a range of average amplitude values. The high rating for Transparency shows that cause and effect concepts were conclusive, despite the gaps when the sphere was invisible. This is well known in Gestaltist psychology and neuroscience: inconsistency does not impede the formation of conclusive cause and effect concepts, or conclusive audiovisual objects, extending Kubovy and Schutz' terminology. Of course the perceived may not correspond to actual mappings, but that is irrelevant to our investigation.

After viewing Interrupted Synchrony, six participants spoke of a single sphere, in spite of the gaps when the sphere was invisible. They grouped the stimuli as a single audiovisual object. This manifests a Gestaltist principle, the one of closure. The principle describes how we "ignore" any intervals so as to infer a single object or form. An example in the visual domain is when we discern a circle in spite the line being interrupted [Wertheimer 1938]. In the auditory domain, an example is when we perceive a sound sequence (e.g. a sound repeating at equal intervals or a note scale) in spite of its actual incompletion, e.g. a missing sound [Snyder 2001]. Closure happened with six participants, but not with the other four. Those spoke of a "small sphere" and a "big sphere", assigning different cause-effect relationships to each. This indicates that they formed two distinct audiovisual objects. It shows that *conclusive cause and effect concepts admit inconsistency and multiple audiovisual objects*.

Combining the mappings from Interrupted Synchrony and Interrupted Random Delay, Complexity produced a global sense of causation, yet the cause-effect relationships were unclear. This shows that *sensing causation does not depend on perceiving how specific changes in the sound relate to specific changes in the image*.

The high rating for Causation means that perceptual binding was convincing. Seven participants assigned multiple visual parameters to multiple sonic parameters, indicating that multiple audiovisual objects were formed at once. Since transparency was rated low, the audiovisual objects remained inconclusive. As a whole, the mapping of Complexity does not conform to pre-existing concepts of causation. While *the synchronised audio-visual components convey causation, the non-synchronised counteract conclusiveness*.

Interestingly, several participants assigned sonic parameters to the visual shape from Interrupted Random Latency. When viewed independently, this mapping had Causation rated low. With Complexity, the participants did possibility not segregate the audio-visual components that produced a sense of causation, from the components that did not. Since Transparency was rated low, they were aware of non-fitting information. Nevertheless, *they sought for a global ecological fit* (Figure 8).

prototype	causation	transparency
Interrupted Synchrony	+	+
Random Latency	-	-
Interrupted Random Latency	-	-
Complexity	+	-

• 1 sphere / single cause-effect - *closure*
• 2 spheres / two cause-effect – no *closure*

• uncertainty / multiple cause-effects
• awareness of non-fitting information
• all components seem in relation

interrupted random delay alone = **NO RELATIONSHIP**
interrupted random delay + synchrony + complexity = **ALL IS IN RELATIONSHIP**

Figure 8: Analysis of the study results

Describing this in short, certain amount of audio-visual synchrony suffices to produce a sense of causation. If complexity generates confusion, we feel unsure about which audio-visual components produce a sense of causation, and which do not. Because we don't understand the cause and effect relationships, the sense of causation extends to the mapping as whole, and we acknowledge sensory information that would have been discarded if concepts were conclusive.

3.6.7 Study Conclusions

A fungible mapping makes us sense causation without distinguishing the base cause and effect relationships. This is characteristic of medium audio-visual fit. The study showed that the fungible mapping includes components that convey a sense of causation and components that do not; it also exhibits complexity so as to confound the actual cause and effect relationships. The study specified that: a) Synchrony conveys a sense of causation even if it exhibits interruptions; one may form a single Gestalt or separate Gestalts, but the cause and effect relationships are conclusive. This means high audio-visual fit. b) When sound and image are mapped with random latency or interrupted random latency, occasional points of synchrony do not suffice to produce a convincing sense of causation. This means low ecological fit. c) Interruptions and diverging interpolations create complexity, confounding which components of the mapping produce a sense of causation, and which do not.

The analysis of results permits further conclusions. Synchrony reveals a cause and effect relationship, but not when points of synchrony are too sparse. Conclusive cause and effect concepts admit inconsistency and multiple audiovisual objects. With inconclusive cause-effect concepts, i.e. inconclusive audiovisual objects, one is aware of non-fitting information, yet perception keeps searching for a global fit.

The aspect of a fungible mapping was gleaned independently from personal creative explorations, so that it can be explored in many different ways and with any audio-visual platform. Synchrony conveys concepts of causation, and interrupted random delay does not. The point is, we do not tell them apart conclusively when the combination of mappings is complex. We are driven to form conclusive concepts at the expense of overlooking or skewing any conflicting information. With a fungible audio-visual relationship, perception continues to acknowledge conflicting information, embracing convergences and divergences as inconclusive concepts.

The study conclusions can be extended to the audio-visual relationship in space, while considering the combination of mappings, physical gesture and physical performance setup. The relation will be fungible if some components create a sense of causation, and others not; the important is that cause-effect relationships are confounded.

3.7 Three Creative Principles for Audio-Visual Instrument Design and Composition

This chapter bridged several areas of research, extrapolating methods to analyse sonic expression and sensory dominance in any audio-visual performance work. These methods were useful to clarify how interfaces can convey a notion of sonic expression, and how complex musical constructions can benefit from the combination with moving images. They enabled me to clarify three creative principles, applicable to audio-visual instrument design, composition and performance:

- **Sound:** to threshold the performer’s control over the instrument and the instrument’s unpredictability, so as to convey sonic complexity and expression. The wealth of sound organisation can be fully experienced when combined with overall visual continuity.

- **Image:** to create overall visual continuity, by applying Gestaltist principles to visual dynamics and dispensing with sudden visual changes, which would automatically attract attention. Visual changeability can be explored at a level of detail.
- **Audio-visual relationship:** to create a fungible audio-visual relationship, combining elements that produce a sense of causation and elements that do not. The combination should exhibit complexity so as to generate inconclusive concepts of causation.

3.8 Chapter Summary and Discussion

The chapter investigated how complex sonic constructions can be fully experienced in audio-visual performance, given a particular notion of sonic expression and the question of visual dominance. It presented methodological tools useful for instrument design, composition and performance. They led to a set of creative principles for the sound, the image and the audio-visual relationship.

Regarding sound organisation, I pinned down a notion of sonic expression related to effort and psychophysical timings. This notion implies that the performer's interaction with the instrument is reciprocal. As Anthony Gritten wrote, "while the subject is certainly performing, it is also performed" [2006:106]. With this he meant that the experience of performing is simultaneously perceived through another type of experience. As a performer I feel that dealing with 'chance' is a way to permeate rather than impose a structure upon the sensory complexity. An instrument is simultaneously a controlled prolongation of the body, and a means of expanding action beyond intention. As such, a threshold exists between its unpredictability and the performer's control. This threshold can be manipulated so as to potentiate one's idiosyncratic expression. An unexpected, often minute event can produce compelling performative tension. It causes a minimal, yet graspable hesitation – a moment of suspense. Resolving the musical challenge in good time then causes a sensation of release. Once musical expression derives from addressing the unexpected resourcefully, performative action must exceed intellectual deliberation because musical timings and intervals possess biophysical logics. Dealing with non-anticipated sonic events makes me acknowledge and respond to sensory details that would otherwise go unnoticed. In a sense, creating musical meaning upon the unexpected augments my sense of control. As my capability of response is challenged, my sensitivity and resourcefulness become greater than if I was strictly executing a plan.

Regarding the image and the audio-visual relationship, I resorted to cognition/ attention research so as to clarify how they can create a surplus of meaning without subordinating audition. Neuroscientific research tells us that sensory dominance depends on attention, and I proposed that intensity derives from the combined effects of automatic and deliberate attention. I defined intensity as the neural impact of any change in the chain of stimuli causing an increase in neural activity. In this way, intensity depends on the event itself, on the stimuli panorama, and on the current perceptual resolution. I proposed a taxonomy that relates attention and intensity with different types of continuity and discontinuity. The terminology distinguishes whether apprehensions are primarily driven through stimuli, or if they are more under individual control. The taxonomy can be used to analyse any sonic and visual dynamics, and compare their relative strength; radical visual discontinuities must be avoided if one desires to foster the experience of complex sound organisation.

The taxonomy is equally useful to analyse audio-visual fit with respect to attention. Synchrony activates concepts of causation, which can be conclusive or inconclusive. Concepts can be conclusive even when inconsistent information reaches conscious awareness. Conclusiveness creates perceptual continuity, as attention prioritises fitting information over non-fitting. I proposed a corresponding notion of high ecological fit. High fit is not demanding in terms of cognitive processing, but attention is likely to focus on the audio-visual composite, rather than on the relations between the sounds. I proposed that medium fit conveys inconclusive concepts, and introduced the notion of a fungible audio-visual relationship. The fungible relationship conveys perceptual binding, but it also creates a level of interpretative discontinuity that loosens the perceptual hierarchy. It conveys a sense of causation but confounds the cause-effect relationships, so that attention embraces fitting and non-fitting information with inconclusive concepts. This was demonstrated with a study, which also exemplified my notions of low and high ecological fit.

The research led me to find three creative principles. The first is to threshold the performer’s control over the instrument, and the unpredictability of sonic outcomes - so that the instrument affords sonic complexity, in a way that suits the performer’s idiosyncratic expression. The second principle is to dispense with radical visual discontinuities. We can explore a wealth of visual changes as long as Gestaltist principles apply to visual dynamics, so as to minimise the demand for visual information processing. And the third principle is to create a fungible audio-visual relationship, which produces inconclusive concepts of causation. The fungible mapping combines synchronised and non-synchronised components, exhibiting complexity so as to appear confusing. This should equally be applicable to the audio-visual relationship in space.

These three creative principles can be parameterised in terms of real-time effort in sound organisation, sonic and visual discontinuities, and ecological fit. Whilst they suit a particular artistic sensibility, the perceptual approach they are driven from should be useful to any audio-visual practice.

CHAPTER 4

A Parametric Model on Audio-Visual Performance

4.1 Chapter Introduction

Instruments and performance situations can be analysed through different parameters, and these can be represented graphically, with a set of axes. Given their abstraction level, parametric visualisation models are useful tools for examining an instrument and the physical performance situation in terms of a general framework of theoretical and practical design decisions. This chapter endeavours to create a parametric visualisation model capable of representing how any audio-visual performance might converge and diverge from the three creative principles described in the previous chapter: sonic complexity, visual continuity, and audio-visual fungibility. In addition, the model should be capable of representing how the work under discussion informs the audience's sense of spatial presence.

Section 4.2 describes three parametric visualisation models, focused on human-computer interaction, on the epistemic dimension of human-computer interaction, and on electronic music performance as audio-visual expression. I examine how each can contribute to the model of the thesis.

Section 4.3 defines the parameters and the graphical aspect of the model. It describes the model in detail.

Section 4.4 uses the model to analyse three audio-visual performances, by Steina Vasulka, Ryoji Ikeda and Thor Magnusson. The investigation exposes how they converge and/ or diverge from each of the creative principles described in the previous chapter.

4.2 Useful Parametric Visualisation Models

My purpose in this chapter is to create a parametric visualisation model capable of indicating sonic expression, sensory dominance and spatial presence. This section presents three related models, outlining how each can contribute to this endeavour.

4.2.1 A Model on Human-Computer Interaction

Birnbaum et al. proposed a parametric visualisation model focused on human-computer interaction, which considered questions about design, options of how to address these questions, and criteria regarding the suitability of the available options [2005]. The model consists of seven axes corresponding to parameters onto which musical devices can be mapped (Figure 9).

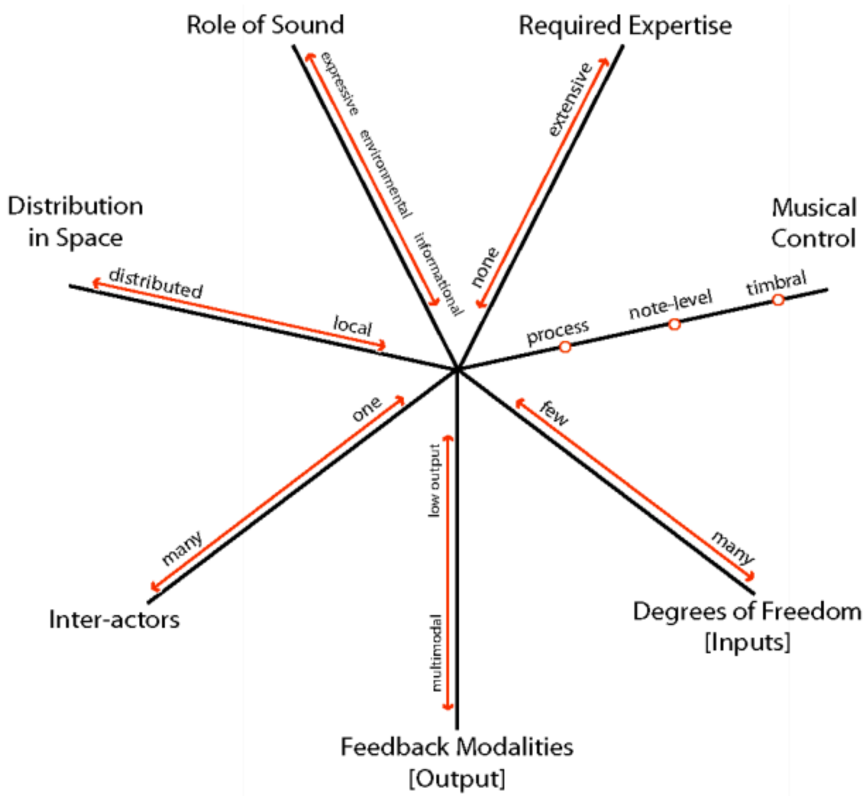


Figure 9: Parametric model proposed by Birnbaum et al., 2005

The *Musical Control* axis contains three discrete points, corresponding to three levels of control over musical processes: timbral, note or score. The parameter does not consider other possibly important levels of control, such as timing. All the other axes in this model are continuous. The axis *Role of Sound* represents the semantic load of sounds; the authors seem to adopt Jeff Pressing's characterisation of sounds, whereby a sound can be *expressive, informational,* or *environmental* [Pressing 1997]. Furthermore, the axis *Distribution in Space* indicates the total physical area inhabited by the instrument. The axis *Inter-Actors* shows the number of people involved in the performance of the device. The axis *Feedback Modalities* represents the degree to which the real-time feedback to the user is multimodal. The axis *Required Expertise* shows the level of practice and familiarity with the system needed in performance. And the axis *Degrees of Freedom* represents the number of input parameters that can be controlled. It does not consider other possible factors determining the degree of freedom, such as digital constrains.

The model from Birnbaum et al. does not consider any visual projection or audio-visual relationship in space, and a few parameters are out of the scope of this research – e.g. variable number of inputs, feedback modalities, or persons operating the musical device. Nevertheless, the model is important to this research for several reasons. It considers sound organisation from the perspective of interaction, as it parameterises the performer's control over the device. This is important to the principle of sonic complexity. The model also considers physical distribution in space, which is important to the notion of performative arena. Moreover, the model adopts an appropriate semantic characterisation of sounds, which can be easily extended to the audio-visual domain. Expressive, environmental and informational semantics convey different modes of spatial presence, and these three semantic dimensions can be parameterised with respect to any audio-visual performance work.

4.2.2 A Model on the Epistemic Dimension of Human-Computer Interaction

Drawing from Birnbaum, Magnusson proposed an eight-parameter model emphasising the epistemic dimension of digital music devices [2010], as shown in Figure 10 (please see next page). Magnusson takes advantage of the star-like graphics so as to visualise oppositions between the parameters.

The *Expressive Constraints* axis represents the amount of digital constraints, which conditions the range of musical possibilities. Opposed is the *Required Foreknowledge* axis, which seems equivalent to the parameter *Required Expertise* from the model of Birnbaum et al. [2005]. Magnusson stresses that many digital systems do not require much musical knowledge in performance as they contain it already. The *Autonomy* axis represents "the degree to which the instrument provides the functionality of an automata" [2010]. In opposition, the *Improvisation* axis indicates how open the instrument is for real time changes, and how quick it can adapt to musical situations. The *Music Theory* axis represents the amount of culturally specific music theory encapsulated in the instrument, in terms of tonal and rhythmical structures, as well as signal processing. A high degree of *Music Theory* seems to imply a large amount of *Autonomy.* But the inverse is not necessarily true, since the encapsulated knowledge may not be characteristic of any culture in particular. The opposed axis, *Generality,* represents the instrument's versatility: "how open in expression" it is and "how well it copes with the multiplicity of musical situations" [2010]. The *Explorability* axis represents how much depth the instrument holds, and how engaging the instrument is. In opposition, the *Creative-Simulation* axis shows whether the instrument is novel in terms of interaction, sound and function, or an imitation of established tools and practices.

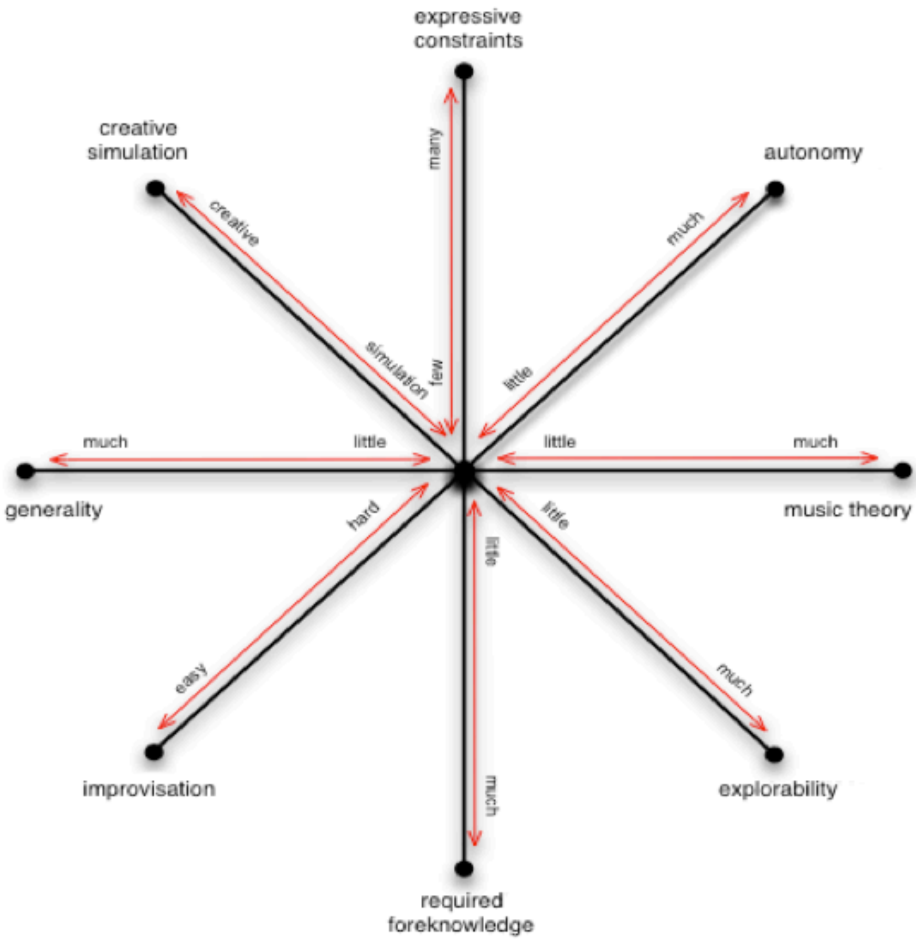


Figure 10: Parametric model proposed by Magnusson, 2010

Magnusson uses the model to analyse a set of different musical interfaces. He stresses that with this model one needs to distinguish between the interface or musical system, and their instantiation as an expressive musical system. For example, he provides separate analyses for Waisvisz' *Hands*[48]. About the *Hands* as a gestural interface he writes:

> Open for improvisation; it affords large areas for exploration and there is relatively high degree of musical foreknowledge required to design the tool or compose for it. The controller is very general (…). From its nature as a sensor device, we see there is little music theory inscribed in it (…), and few expressive constraints. [2010]

About the *Hands* as a musical instrument, he writes:

> The music theory has become stronger, there is more autonomy and expressive constraints have been designed into the system. It is therefore less general and can be used in fewer musical contexts. It can still be ideal for improvisation. [2010]

[48] http://www.crackle.org/TheHands.htm

Magnusson's model is expressive in representing how digital music devices condition human-computer interaction, and that is important to the practical work of the thesis, which explores disparities between acoustic and digital outputs as creative material. Furthermore, the parameters *Expressive Constraints, Explorability, Autonomy, Music Theory* and *Improvisation* relate to the principle of sonic complexity. The model will be useful to discuss certain similarities and differences amongst the iterative versions of the audio-visual software developed in this research. By extension, it will be useful to discuss the model presented in this thesis, which aims to visualise the common thread between those iterations. Furthermore, the idea that a parametric visualisation model can be used to analyse a same work from different perspectives will be useful when discussing which aspects in my practical work are versatile, and which are not.

4.2.3 A Model on Electronic Music Performance as Audio-Visual Expression

Ciciliani proposed a model to represent electronic music performance as an audiovisual means of expression [2014]. The model seeks to distinguish what the author calls *centripetal* and *centrifugal* performance tendencies (Figures 11a and 11b, please see next page). In centripetal performances, the performer is visible and the centre of attention. The relation between his physical action and the sonic results is clearly perceivable, and the sound sources are placed near to him. In centrifugal performances, the performer is in a hidden position. He functions as a controlling rather than enacting entity; there is little or no correspondence between his physical actions and the sonic results, and sound sources are spread out in space.

The model includes four parameters characteristic of centripetal performances. The *Body* axis shows whether the performer's body is clearly visible, and the *Presence* axis indicates whether the performer's presence is prominent or not. The *Embodiment* axis indicates whether there is a strong correlation between the performer's bodily actions and the sonic results. Importantly, Ciciliani distinguishes between technological and visible correlations: his *Transparency* axis indicates if there is a strong readability between the performer's actions and the sonic result.

Ciciliani identifies three parameters characteristic of centrifugal performances (later he adds a fourth). The *Space* axis shows whether the sound source is near the performer or spread through the performance space. This parameter seems equal to the *Distribution* axis in the model from Birnbaum et al. [2005]. The *Mediatization* axis shows to which extent sounds occur independently from the performer's physical actions. We can say that Magnusson's *Autonomy* parameter [2010] implies a high level of *Mediatization*. The *Camouflage* axis shows whether the performer's actions are hidden from the audience. In the first paper about this model [2014], Ceciliani described the *Camouflage* parameter as the inverse of *Transparency*. Interestingly, the subsequent user study reported in [Ciciliani and Mojzysz 2015] showed that the evaluation of *Transparency* is more variable than the evaluation of *Camouflage*.

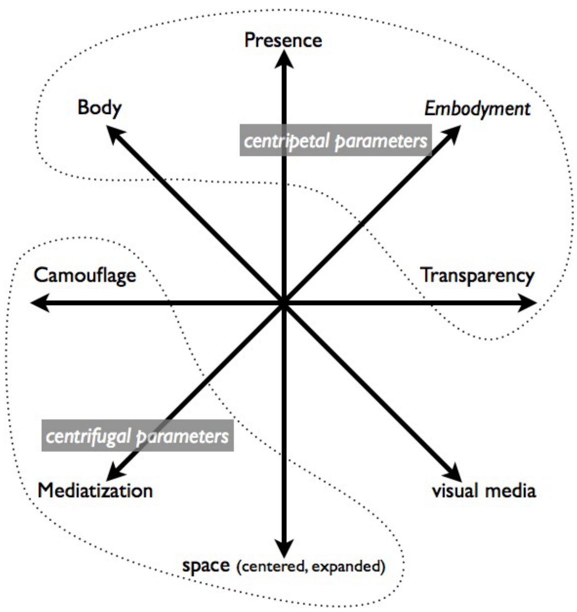


Figure 11a: Parametric model proposed by Ciciliani, 2014

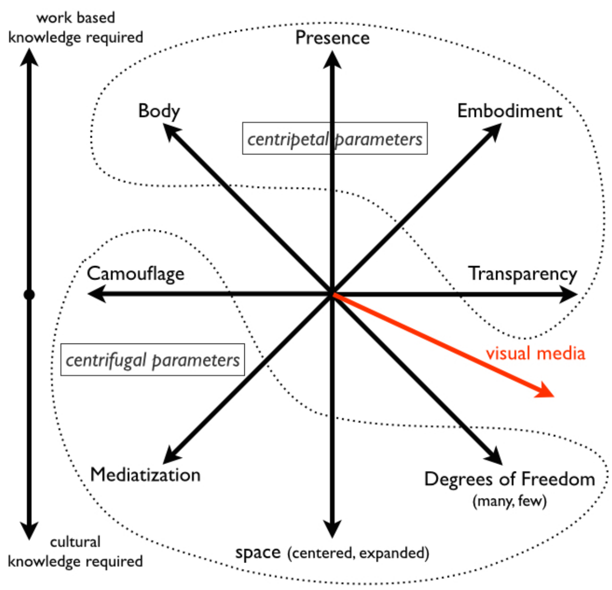


Figure 11b: Second iteration of the model, by Ciciliani and Mojzysz, 2015.
Note: the Visual Media parameter was subsequently discarded.

Ciciliani also attempted to include a floating axis, called *Visual Media*. Depending on the work in question, it could be classified as centripetal or centrifugal. It would represent all sorts of visual extensions, like video projections or sound objects. The study from [Ciciliani and Mojzysz 2015] led him to discard the parameter, as the model could not indicate the function of these extensions. At this later research stage, he also added two parameters to the previous model. Adapted from Birnbaum's model [2005], *Degrees of Freedom* indicates the number of expressive parameters that the performer can control in real-time. The second new parameter, placed independently from the star-like graphic, represents the *Work-Based Knowledge* and the *Cultural Knowledge* required to perform the device. In the study from [Ciciliani and Mojzysz 2015], people found it hard to distinguish one from the other. *Work-Based Knowledge* seems equivalent to the parameter *Required Expertise* from [Birnbaum et al. 2007], and *Cultural Knowledge* seems equivalent to the parameter *Required Foreknowledge* from [Magnusson 2010].

Ciciliani's model is significant to this research due its concern with the audience's experience in space. The *Embodiment* parameter concerns the performer's mode of interaction with the instrument; this is important to the principle of sonic complexity, but the term embodiment has many diverging implications, as noted in chapter 2.5.1. The *Transparency* parameter indicates if there is a strong readability between the performer's actions and the sonic result. The model of this thesis needs to be more specific with respect to "readability", since the distinction between sense of causation and perceived cause and effect relationships is crucial to sensory dominance and to the principle of fungible audio-visual relationships. Furthermore, Ciciliani's *Body* parameter shows whether the performer is visible or not. That is equally relevant to the model of this thesis, which needs to indicate whether the audio-visual relationship involves physical gesture. And last but not least, Ciciliani's notions of "centrifugal" and "centripetal" performance are extremely useful. "Centrifugal" denotes a focus on the performer's expressivity, and "centripetal" denotes a focus on the environment. One can look at how these notions are informed by Ciciliani's parameters, so as to clarify the surplus of creating separate parameters for semantic characterisation. In fact, the audience's sense of spatial presence depends on the physical setup and the instrument design, but it also depends on the specific qualities of sound and image.

4.3 A Parametric Model on Sensory Dominance, Sonic Expression and Spatial Presence

This section presents a parametric model on audio-visual performance, which aims to parameterise the three creative principles described in the previous chapter, as well as the physical setup and the global semantics of any work.

4.3.1 Learning from Problems: General Considerations about Parameters and Graphical Aspect

The models described in the previous section present several problems relatively to the aims of this research, and we should find ways to circumvent those problems.

As Ciciliani attempted to represent "the function of any visual extension", his investigation made it obvious that such function cannot be summarised with a single, floating parameter [2014, 1015]. Any new model should not attempt to parameterise experience itself, nor should it attempt to parameterise all details relevant to the work. Whilst the former would be too ambivalent with respect to terminology, the latter would be too specific with respect to technical means and aesthetical approaches.

In contrast with the models from Birnbaum and Magnusson [2005, 2010], the model of this research seeks to be applicable to any audio-visual instrument or system, regardless of whether it is electronic, digital, acoustic, or any combination of the previous. This requires a much greater level of abstraction with respect to interaction than those models provide. That greater level of abstraction also conveys the aim of parameterising audio-visual performance/ composition, which requires many parameters beyond interaction.

Star-like graphics can serve the purpose of emphasising oppositions, as happens in the models from Magnusson and Ciciliani [2010, 2014]. Synthesising information is then not a priority: the eight opposing parameters from these models could also be represented with four axes. Star like graphics seem less useful when the parameters are not in opposition; particularly when some axes contain discrete points, as happens in Birnbaum's model. Birnbaum's parameters are still legible because the model only includes seven axes; yet they seem easier to read in the orthogonal graphic of Figure 12.

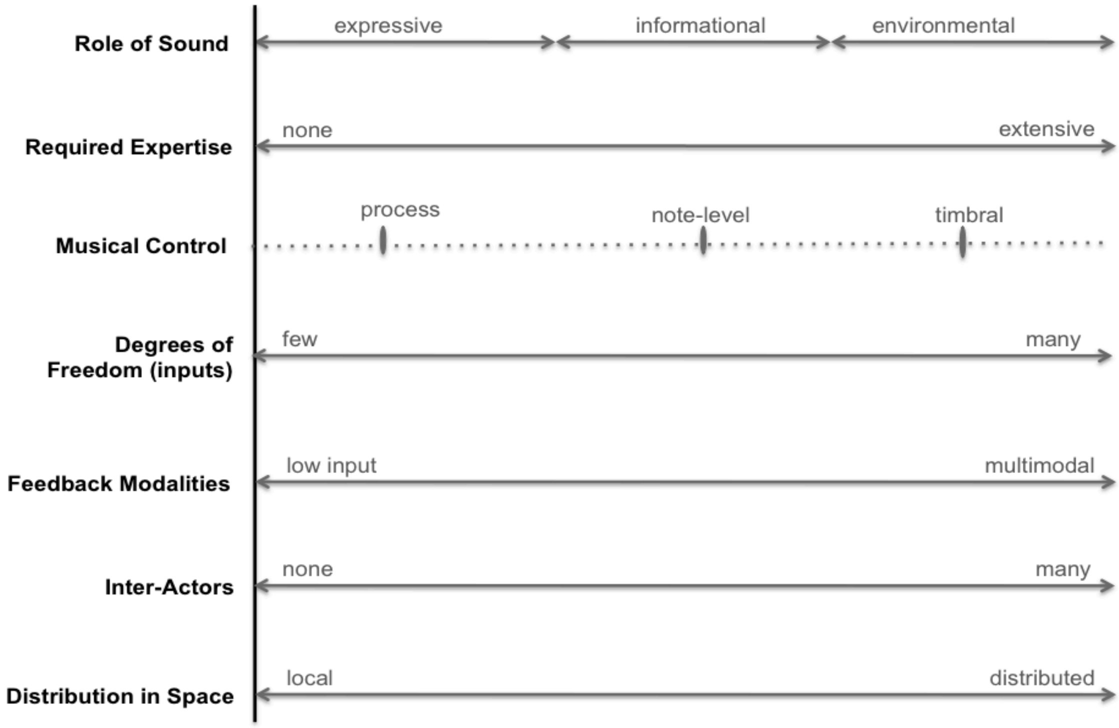


Figure 12: Birnbaum's parameters represented in a Cartesian graphic, rather than a star-like graphic

The model of the thesis will use orthogonal rather than star-like graphics, because we do not want to subordinate the selection of parameters to the intent of creating oppositions, and horizontal text is easier to read.

Our parametric model should dispense with any conceptual separation between instrument, composition and performance, so as to facilitate integrated analyses. The model should be useful to discuss sonic expression, sensory dominance and spatial presence in any audio-visual performance, including the practical work developed in this research. Beyond its general applicability, our model should also consider how to represent the creative principles explored in this research: a) sonic complexity, entailing thresholds between control and unpredictability; b) visual continuity, informed by Gestaltist principles; and c) audio-visual fungibility, which creates a sense of causation, and simultaneously confounds the cause and effect relationships. Furthermore, the model should represent the physical setup and the semantics of the work, as these are important for one's sense of spatial presence.

4.3.2 Parameterising Sonic Expression & the Principle of Sonic Complexity

Chapter 2.5 grounded the present notion of musical expression, questioning how it differs from other notions. The distinguishing factor is the amount of real-time effort in sound organisation. Chapter 3.3 presented a method to analyse how effort informs interaction design and musical expression. I defined the creative principle of sonic complexity, where thresholds between control and unpredictability make the interaction somewhat effortful. Whilst control conveys embodiment and performatory gestures, unpredictability creates a separation that conveys exploratory gestures. This conveys sonic constructions that combine a hierarchical level of complexity and a dynamic level of complexity (using Snyder's terminology [2001], which has also been related with Ryan's, in chapter 3.3.2).

In summary, sonic expression depends on the performer's interaction with the instrument/ system, on the sonic dynamics, and the semantics of sound. Figure 13 shows the relevant variables.

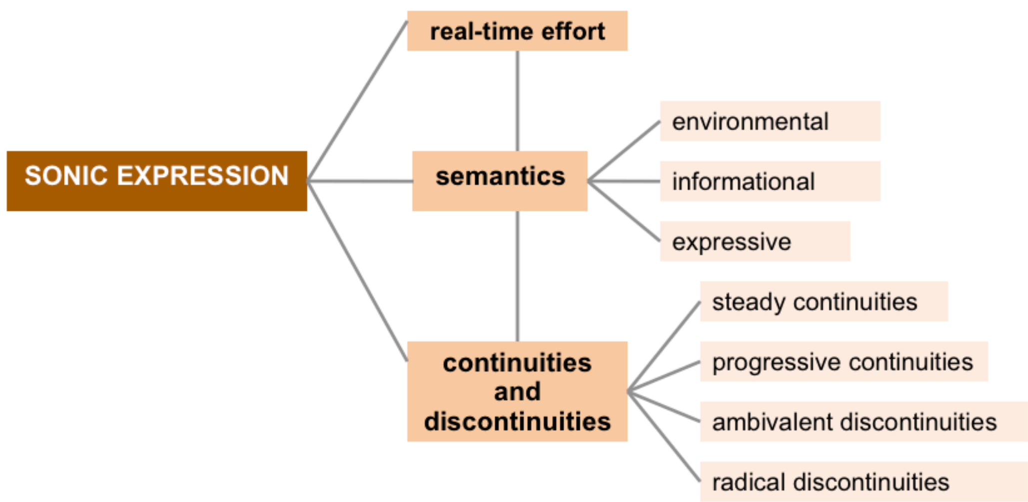


Figure 13: Factors that influence sonic expression

To assess on whether the work converges or diverges from the creative principle of sonic complexity, an axis for real-time effort in sound organisation will suffice: both little effort and extremely high effort diverge from principle. As previously explained, this research considers the notion of effort relatively to cognitive demand and musical timing. Little effort can mean that the relation between deliberate human agency and sonic results is linear and clearly perceivable. Alternatively, it means that the music does not depend much on the performer's interaction. In any case, it indicates a divergence from the principle of sonic complexity. Medium effort indicates a convergence with the principle. It denotes that the instrument requires particular skills, yet a sense of immediacy conveys musical timing, and/ or technical configurations rule out undesired outcomes. Furthermore, a high level of effort implies particular skills and/ or high cognitive demand; the interaction with the system does not feel immediate, and/or the system does not rule out any undesired outcomes. Medium-high effort can indicate a convergence with the principle of sonic complexity, as long as the challenge of creating local time (using Ryan's terminology [Sa et al. 2015]) is proportional to the musician's mastery. Maximum effort diverges from the principle, as the performer does not have any control over the sonic construction.

As a parameter, real-time effort seems at an appropriate level of abstraction. It is applicable to both digital and acoustic instruments – it can reflect the skills required for live coding, as well as the skills for playing an acoustic instrument. Furthermore, it summarises information represented in the models of Birnbaum and Magnusson. The amount of real-time effort can result from the combination of many factors, for example Birnbaum's *Musical Control*, *Degrees of Freedom* and *Feedback Modalities* [2005], and Magnusson's *Expressive Constraints, Explorability, Autonomy, Music Theory* and *Improvisation* [2009]. Also, the parameter is directly proportional to *Required Expertise* [Birnbaum et al. 2005] and *Required Foreknowledge* [Magnusson 2009]: the more effortful is the instrument, the more expertise and practice is required for its performance.

Our notion of sonic expression embraces a diversity of musical idioms, with different types of continuities/ discontinuities, and different semantic dimensions. Hence, visualising dynamics and semantics does not enable us to infer whether the work under discussion converges or diverges with the principle of sonic complexity. Nevertheless, these parameters are certainly important when discussing the sonic construction, and the work as a whole. It is advantageous to use Pressing's terminology [1997] for parameterising semantics: the semantics of sound can be informational, expressive and environmental. And to parameterise sonic dynamics, we should use the taxonomy presented in chapter 3.4, which relates continuities and discontinuities with attention.

As explained, perceived discontinuities depend on stimuli, panorama and perceptual resolution. The panorama also incorporates time, but the relation of time and attention is not linear. For example, a long period of continuity can make an inconsistent event more discontinuous, but that is not a given because continuity also leaves attention under greater individual control. Sustained attention would increase perceptual resolution, making the discontinuity more intense. Without attention, the inconsistency would be less intense; it could even go unnoticed.

The construction of musical time can be very complex, and an additional axis for time would not be useful in our model because a single graphic representation is not capable to indicate how attention changes over time. As long as we explicitly refer the timescale of the represented continuities/ discontinuities, the model can be used to visualise moment-by-moment situations or summarise large compositional structures.

4.3.3 Parameterising Sensory Dominance & the Principles of Visual Continuity and Fungible Audio-Visual Relationships

Sections 3.4, 3.5 and 3.6 of chapter 3 showed that we can analyse sensory dominance based on visual dynamics and ecological fit. The diagram in Figure 14 shows the relevant variables.

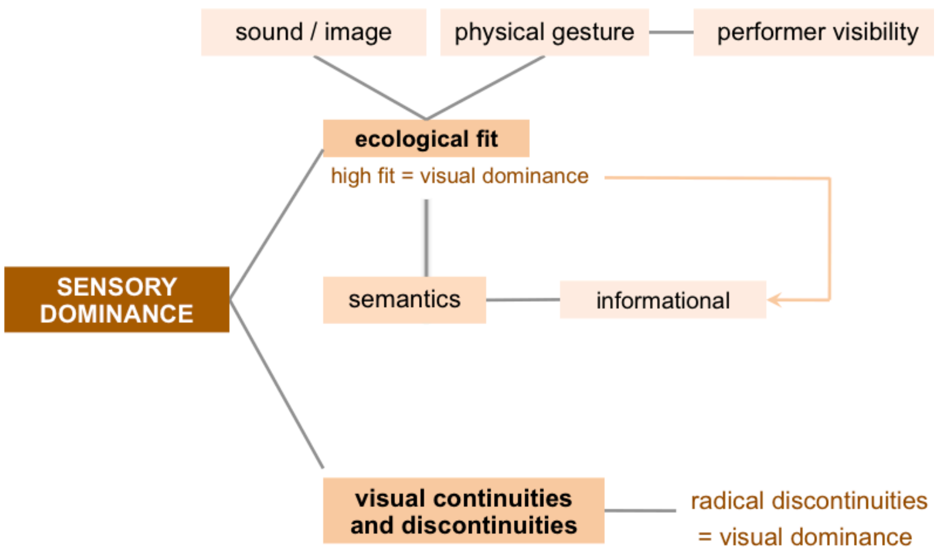


Figure 14: Factors that influence sensory dominance

Visual dynamics should be parameterised using my taxonomy of continuities/ discontinuities (chapter 3.4.2). Radical visual discontinuities attract automatic attention [e.g. Knudsen 2007], causing vision to dominate over audition in multisensory integration [e.g.

Kobayashi 2007]. For example, the ventriloquist effect (i.e. the perceptual dislocation of the sound source toward the visual target) occurs with automatic attention, but not with deliberate attention [Bertelson et al. 2000].

Dispensing with radical visual discontinuities is complying with the creative principle of visual continuity. Some discontinuities might be more difficult to assess then others: there is a gray area, which respects to how time and individual predisposition affect the interplay of deliberate and automatic attention. Indeed, the threshold between ambivalent and radical discontinuities depends on more than the visual behaviours of an audio-visual system. If over time one keeps paying attention to the visual details they become gradually more intense, and one becomes more susceptible to automatic visual attention. However, it is not mandatory that one keeps attending the image. Directing the eyes to a visual target (the screen) increases visual resolution by default [Knudsen 2007], but solely at first. Everyone knows by experience that over time, we can “look without seeing”, just like we can “hear without listening”. Representing individual predisposition is not useful in our model, but we can definitely indicate the timescale of experience so as to better contextualise the perceived continuities and discontinuities.

There could be another gray zone here. Even without radical discontinuities, if the visual dynamics are more discontinuous than the sonic construction, attention might prioritise discontinuity. Yet those visual discontinuities will not lead to automatic and unconscious multisensory integration processes, and one can also deliberately pay attention to the sonic construction.

We can parameterise ecological fit based on our distinction between conclusive and inconclusive concepts of causation (chapter 3.5) and our study about audio-visual mapping and perception (chapter 3.6). An audio-visual relationship conveying conclusions about causes and effects exhibits high ecological fit, even if the cause-effects are not consistent (as happens with inconsistent synchrony). Conclusive concepts of causation make perception prioritise fitting information over non-fitting, and disregard the relation between the sounds. It is possible to invert perceptual prioritisation when concepts are inconclusive, but there will be no convincing sense of causation with low ecological fit (as happens with interrupted random delay). Then, perception can bind separate representations of sound and image, but that is demanding in terms of cognitive processing. Thus, attention is likely to focus on making sense of the audio-visual relationship, rather than on the music.

A work solely converges with the creative principle of audio-visual fungibility if the audio-visual relationship exhibits a medium level of ecological fit, so that one forms causal concepts, but these remain inconclusive. The study from chapter 3.6 showed that fungible relationships combine components that convey a sense of causation, and components that do not; there must be complexity enough to confound the cause-effect relationships. The same is applicable to the fungible audio-visual relationship in space.

To parameterise ecological fit we should consider three factors. The first is that our assessments about ecological fit are based on concepts of causation, because understanding causation in good time is a primary aim of the brain (chapters 2.4 and 3.5). Hence, the model does not require a separate axis for causation. The second factor is that we must take into account the audio-visual mapping as well as the audio-visual relationship in space, including physical gesture, if visible. The third factor is that audio-visual fit entails informational semantics because it makes us infer causation, but we cannot use a single axis to visualise fit and informational semantics, because representative sounds and images entail that type of semantics as well.

In summary, to visualise sensory dominance the model requires an axis for visual discontinuities, and two axes for ecological fit; one for the relation between sound and image, the other for the relation between physical gesture and system output. An axis for the performer's visibility can show whether the perceived audio-visual relationship embraces physical gesture. We can infer that a work conveys visual dominance when the model shows radical visual discontinuities, high ecological fit, or both. Audition can be command in any other case.

4.3.4 Parameterising Spatial Presence

Spatial presence corresponds to the audience's experience of the performative arena, using Emmerson's terminology [2007]. The performative arena is inextricably related to semantics, as it embraces the physical and the psychological space of the work. Pressing's characterisation of sounds is useful here; it is equally applicable to the audio-visual domain. Our model considers that spatial presence depends upon the audio-visual relationship, the qualities of sound and image and the physical setup (Figure 15). The product exceeds the sum - the possibility (and desire) of reliable verification. However, the model can provide cues.

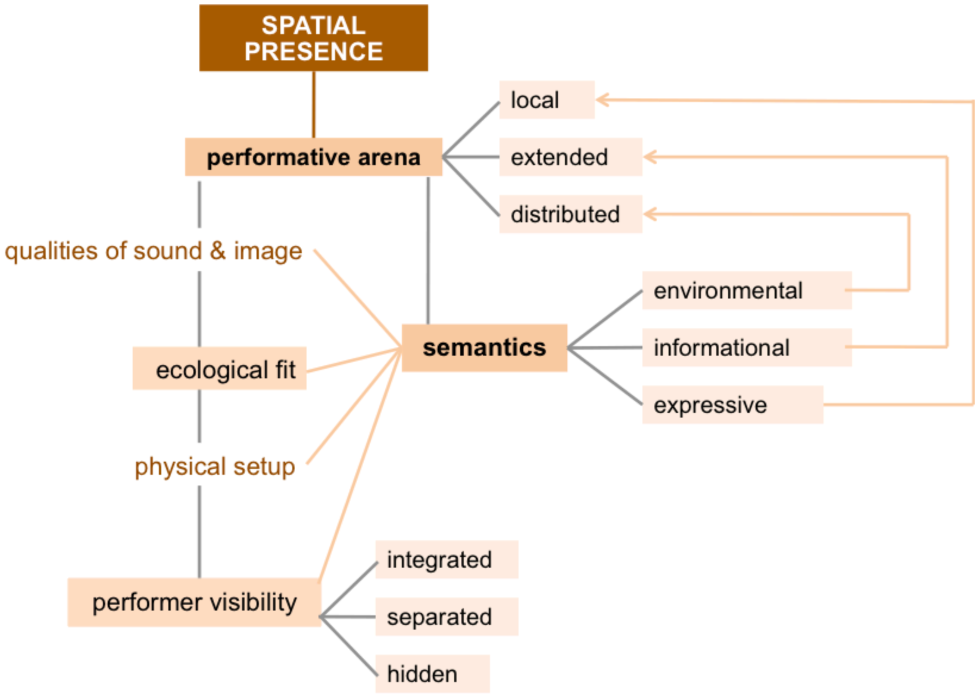


Figure 15: Factors that influence the sense of spatial presence

Extending Ciciliani's understanding of "centrifugal" and "centripetal" performance tendencies [2014], we can characterise the performative arena as local, distributed, or extended. The local arena implies expressive semantics, derived from the qualities of sound and image as well as from the loudspeaker placement and lighting. The distributed arena implies environmental semantics, derived from loudspeaker placement and lighting. And the extended arena implies informational semantics, derived from the qualities of sound and image as well as from the audio-visual fit. Ciciliani's model does not address the possibility of an extended performative arena.

The arena is extended when the sense of spatial presence expands beyond the physical space. At least a certain amount of audio-visual fit is required to convey one's spatial presence in a visible, virtual space beyond the screen [e.g. Dourish 2004, Sacau 2008]. However, we can also be mentally transported when listening to recognisable sounds in darkness, or watching a landscape in silence. In other words, one's mental transportation can be driven by the informational load of the sound or the image, combined with the loudspeaker distribution.

The qualities of sound and image can be parameterised in terms of interaction, dynamics and semantics. We need to parameterise those variables in order to assess on sonic expression and sensory dominance. We parameterise interaction in terms of real-time effort in sound organisation. To parameterise sonic and visual dynamics, we use the taxonomy of continuities and discontinuities. And to parameterise semantics, we draw from Pressing [1997] and Birnbaum et al. [2005].

A parameter for the performer's visibility can indicate the physical relation between the performer and the instrument/ system, by showing whether the performer is integrated, separated or hidden. This relates to the *Distribution in Space* parameter from [Birnbaum et al. 2005], which reflects the total physical area inhabited by the instrument. If the performer is integrated, the image functions as a stage scene.

4.3.5 Parameter Visualisation

The previous subsections summarised how we can infer sensory dominance, sonic expression and spatial presence from a set of variables. The model can articulate information with respect to the performer's interaction with the instrument/ system, the sonic and visual dynamics, the audio-visual relationship, the performative arena, and the semantic dimensions of the work. The challenge here is that the functionality of any model requires a limited amount of parameters; but with an orthogonal graphic we can also create sub-parameters without compromising legibility. Figure 16 shows the model.

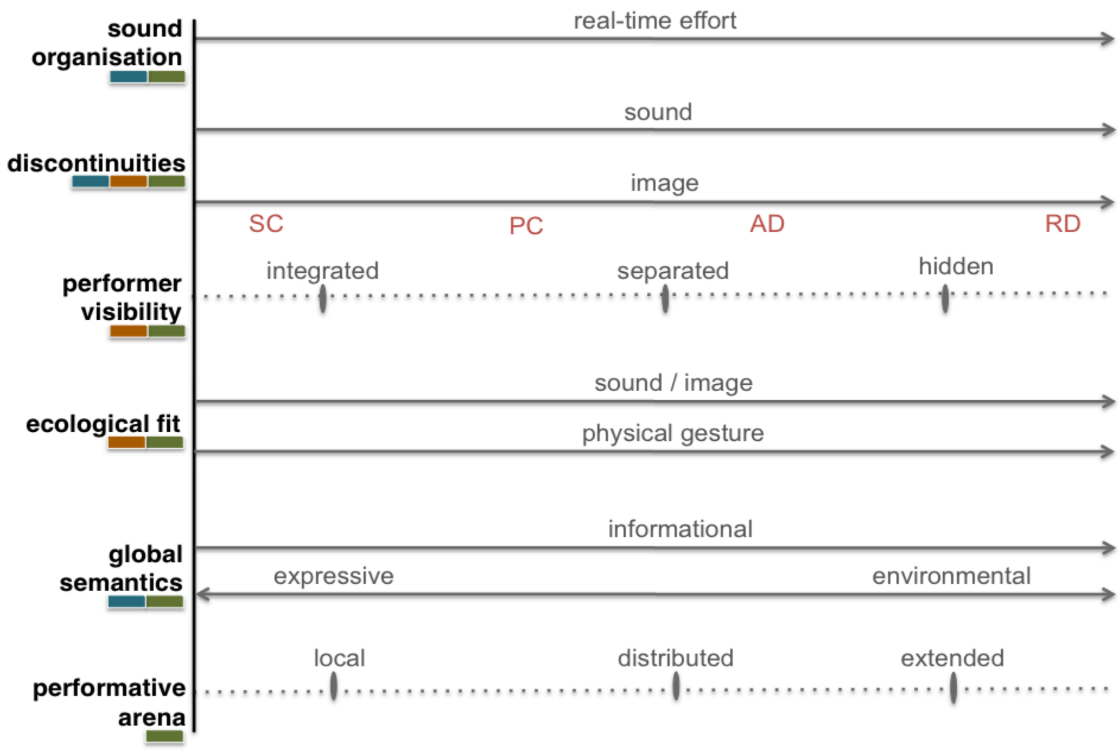


Figure 16: Visualisation of the parameters from which we can infer sonic expression (blue), sensory dominance (orange) and spatial presence (green)

- The *Sound Organisation* parameter represents to which extent the sonic results depend on the performer's *Real-Time Effort*, motor and/ or conceptual. Little effort can mean two things: either the relation between deliberate human agency and sonic results is linear and clearly perceivable, or the music does not depend much on the performer's interaction. Medium effort denotes complex interface behaviours and required expertise, but also a sense of immediacy and/ or technical configurations that rule out undesired outcomes. High effort indicates that the interaction with the system does not feel immediate, and/ or the interaction is demanding in terms of cognitive processing, and/or the system does not rule out any undesired outcomes. Both little effort and very high effort indicate a divergence from the creative principle of sonic complexity.

- The *Discontinuities* parameter includes two sub-parameters: one for the range of *Sonic Discontinuities* and the other for the range of *Visual Discontinuities*. To represent the sonic and visual dynamics we use the taxonomy presented in chapter 3.4, which distinguishes Steady Continuities (SC), Progressive Continuities (PC), Ambivalent Discontinuities (AD) and Radical Discontinuities (RD). A sensory modality exhibiting radical discontinuities prompts automatic attention. A lower level of discontinuity means that deliberate attention is required. Radical visual discontinuities convey visual dominance, indicating that the work diverges from the creative principle of visual continuity.

- The *Performer Visibility* axis shows whether the performer appears *Integrated* or *Separated* from the visual projection, or if she is *Hidden*. The parameter shows to which extent physical gesture influences our assessments on ecological fit. It also provides cues about the scale of the instrument/ system, which influences the sense of spatial presence.

- The *Ecological Fit* parameter shows to which extent what we see fits with what we hear. Low fit means a weak sense of causation. High fit means conclusive concepts of causation. Medium fit means inconclusive concepts, and convergence with the creative principle of fungibility. The parameter includes two sub-parameters: one for the *Fit Between Sound and Image*, and the other for the *Fit Between Physical Gesture and System Output*. In this way, the model can show to which extent our conclusiveness or inconclusiveness derives from the audio-visual mappings, or the audio-visual relationship in space.

- The *Global Semantics* parameter extends Pressing's characterisation of sounds [1997] to the audio-visual domain. The semantics of an audio-visual performance work depend on the semantics of the sound, the image and the audio-visual mapping, on the performer's visible gesture, her location, and the distribution of sound sources in space. The *Informational* dimension of a work prompts causal and semantic percepts. The

Expressive dimension indicates a focus upon the performer's personal expression and skills. The *Environmental* dimension indicates a focus upon space and context.

- The *Performative Arena* parameter represents the relation between the physical and the psychological space of the work. The model distinguishes three types of arena, corresponding to different modes of spatial presence. *Local* denotes a focus upon the performer, corresponding to Ceciliani's centripetal performance tendencies [2014]. It denotes a sound source placed next to the performer, and/ or her physical location next to the visual projection. The local arena relates to the expressive semantics of the work. *Distributed* means that the sound source(s) and/ or the visual projection(s) are distributed in space. It corresponds to Ceciliani's centrifugal performance tendencies [2014]. The distributed arena relates to the environmental dimension of the work. *Extended* means that the sense of spatial presence exceeds the physical performance location. The extended arena denotes the audience's mental transportation to another space. This depends on the informational load of the sound, the image and the audio-visual relationship (regardless of whether the sounds and images are more expressive or environmental). Extending the audience's spatial presence to a virtual world beyond the screen requires at least a medium level of audio-visual fit.

4.4 Analysing Different Audio-Visual Performance Languages

This section demonstrates that the parametric model can be used to analyse sonic expression, sensory dominance and spatial presence in any audio-visual performance. I will discuss how performances by Steina Vasulka, Ryoji Ikeda and Thor Magnusson converge and diverge with the principles of sonic complexity, visual continuity and audio-visual fungibility. My analyses are based on video documentation available online, and I also experienced these works in person.

4.4.1 Performance by Steina Vasulka

The photos and the graphic in Figure 17 (please see next page) represent a performance by Steina Vasulka, integrated in *Strange Music for Nam June Paik*, which took place at the Smithsonian American Art Museum in 2012.

The work pertains to Vasulka's *Violin Power* series, in which she controls video with an electric violin with MIDI output (the *Zeta Violin*). Her website introduces the series[49]. It explains that in 1991, after having experimentally interfaced her acoustic violin with a variable speed video cassette player, she bought a MIDI violin and a Pioneer Disk Player. Interfacing these instruments with a computer gave her "an instant access to any frame of video on the disk as well as access to fast/slow and forward/backward movements". The initial software was written by Russ Gritzo and further developed by Bill Heckel. In the late 1990s she begun using the software Image/ine, developed by Tom Demayer at Steim.

The video of this particular performance is on youtube[50]. It lasts 37"27". Vasulka stands in darkness, next to the video projection. The video source material is processed in real-time. It includes a shot of trees shaking in the wind, several shots of Vasulka playing the violin, and a shot of Waisvisz playing *the hands.* The sonic construction combines the sound of the violin and the sound of the video, digitally processed.

[49] http://www.vasulka.org/Steina/Steina_ViolinPower/ViolinPower.html
[50] https://www.youtube.com/watch?v=9mg1weOHWSs

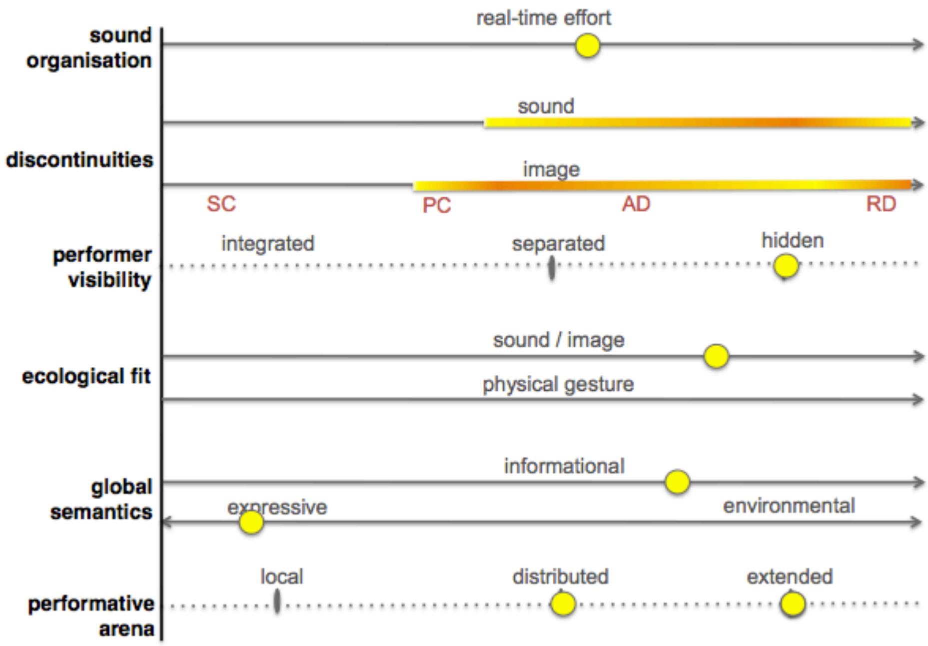


Figure 17: Steina Vasulka's "Violin Power" performance at Strange Music for Nam June Paik, 37'27", https://www.youtube.com/watch?v=9mg1weOHWSs

Real-Time Effort in Sound Organisation

The violin strings, tuned to standard A, C, D, E and G, are assigned to framing locations on the hard drive, controlling the speed and direction of video playback, and addressing segments on the disk. There is obviously something in common between the Zeta Violin and the instrument developed in this research: the interaction involves coordinating the immediate output of a string instrument and mediated, digital outputs. But Vasulka's interaction is more effortful: she needs to coordinate real-time decisions about the sonic and the visual outputs, whereas I focus solely on the sonic construction. The mastery of her work reflects her decades-long investment in playing and developing the instrument.

Discontinuities

Vasulka intertwines the violin sound and the video sound so as to create a fair amount of sonic discontinuities. The visual dynamics change in clearer ways than the sonic dynamics. We can say that they structure the work, for two reasons:

a) The transition from one video segment to another is often very sudden, creating radical visual discontinuities. Examples:
   - At 3'55'' there is a sudden change from trees shaking in the wind, to old footage of Vasulka playing the violin.
   - At 5' there is another sudden change to a close-up of the violin, digitally transformed.
   - At 6'40 the image changes suddenly to Waisvisz playing *the hands.*

b) In a same video segment, normally the visual dynamics convey Gestalts of invariance, providing a sense of visual continuity. Examples:
   - 2'1"-3'55" (trees shaking in the wind)
   - 3'55"-5' (Vasulka playing the violin)
   - 5'-6'40'' (close up of someone playing the violin, with and without digital processing)
   - 6'40-12'5" (Waisvisz playing *the hands*, with and without digital processing).

   In this last section, the sense of continuity proceeds when Vasulka starts processing the image – the image conveys the Gestaltist principle of invariance. There are sudden auditory changes (e.g. at 7'33") but they do not really disrupt our sense of continuity. Sudden/ synchronised audio-visual changes are disruptive though, as explained in chapter 2.4.2. Hence, although Vasulka uses the same video material from 0'40" to 1'11", we can distinguish two sections in this time frame. The transition happens at 1', because a sudden sound synchronises with a sudden visual change, and there is a significant change in the digital processing of the image.

Performer Visibility

The performer is on stage. We see the setup when she starts playing, but as lights turn off we solely see the video projection.

Ecological Fit

The ecological fit is assessed based on the qualities of the sound, the image and the audio-visual mapping; Vasulka's physical gestures are not visible, and the loudspeaker placement is secondary to the audio-visual relationship because the video output is mono. There is great ecological fit in this work, because:

a) Vasulka utilises video sound, and there is a high level of audio-visual fit intrinsic to the video footage itself. From 2'1"-3'55" we hear the visible trees shaking in the wind; from 3'55"-5' we hear the young Vasulka playing the violin; from 5'-6'40'' we hear the violin that we see in close up; from 6'40-12'5" we hear Waisvisz playing *the hands.*

b) The violin sounds seem to affect the video in clearly perceivable ways. The direction of causation is clear, as the sounds of the violin and the video are clearly distinguishable; Vasulka plays the violin with intervals, leaving space for the video sound to unfold. For example, from 2'03" – 2'30" (when we see and hear the video of the trees) the continuity of the image matches the continuity of the sound. When the violin sound appears at 2'30", both video sound and image start morphing accordingly.

Global Semantics

The semantics of this work are fundamentally informational and expressive. The informational dimension derives from two factors. One is the recognisable violin timbre - we do not see physical gestures, but we understand when the strings are touched. The other is the ecological fit between the video image and the video sound. The informational dimension reinforces the expressive: the recognisable violin timbre and the ecological fit lead us to focus on Vasulka's skills and expressiveness, although she is not visible.

Performative Arena

The audio-visual fit extends the performative arena beyond the physical performance room, as the video footage reports to different situations and spaces. There are loudspeakers placed around the audience, hence the performative arena is also distributed. But the arena does not entail a local dimension, despite the expressive semantics of the work: Vasulka is hidden in darkness.

Based on these parameterisations, we can assess how the work converges and diverges with the creative principles of sonic complexity, visual continuity and audio-visual fungibility. The axis *Real-Time Effort in Sound Organisation* indicates convergence with the principle of sonic complexity. The *Visual Discontinuities* axis shows certain amount of radical discontinuities, implying that the work diverges from the principle of visual continuity. The *Ecological Fit* axes show that cause and effect relationships are quite clear, which makes the work diverge from the principle of fungibility. For these reasons, we can say that vision dominates over audition.

4.4.2 Performance by Ryoji Ikeda

The photos and graphic in Figure 18 represent a performance piece conceived and directed by Ryoji Ikeda. The piece is called *Superposition*, and the video is 1'37'' long.

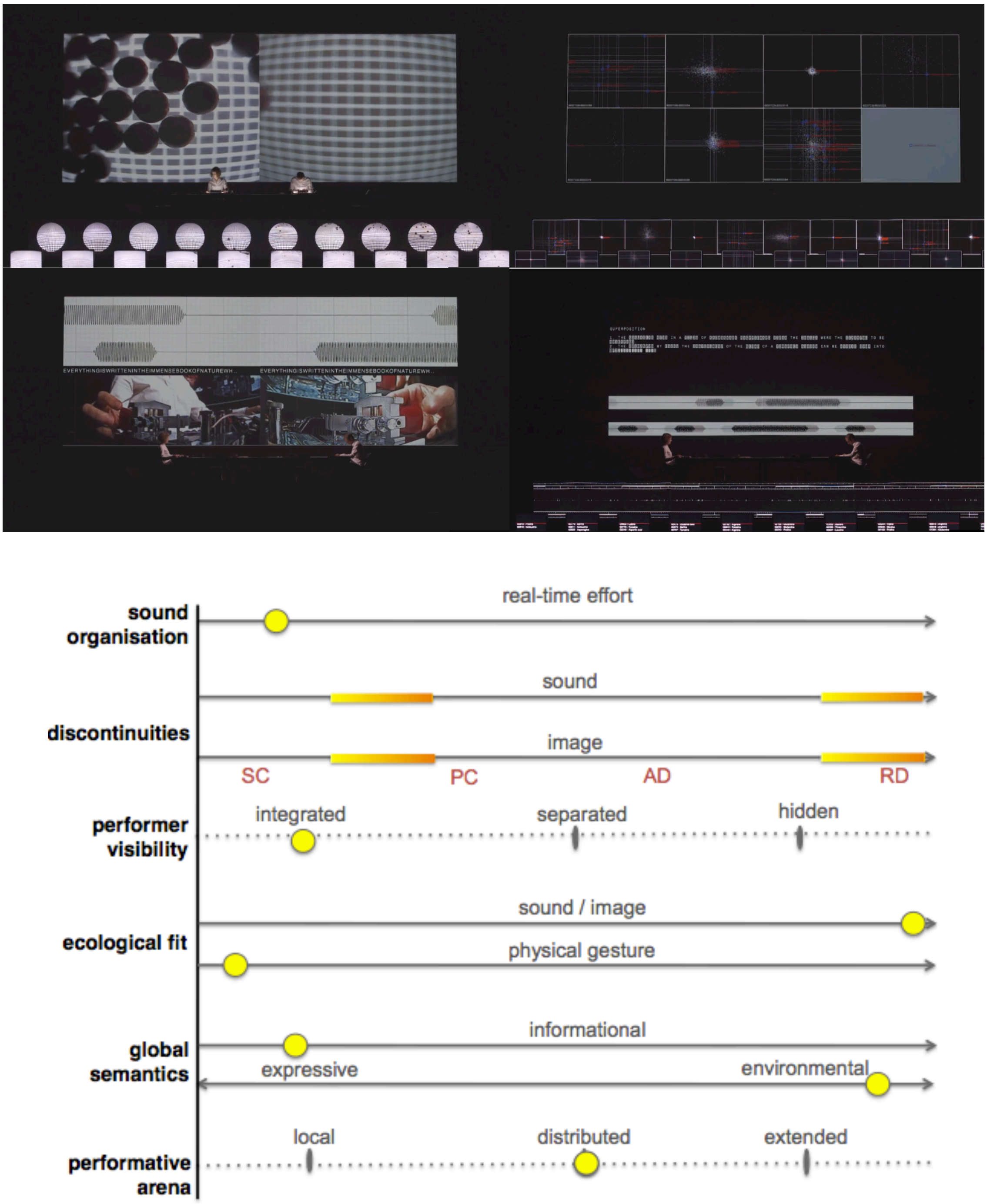


Figure 18: Ryoji Ikeda's performance "Superposition", 1'37"
http://www.ryojiikeda.com/project/superposition

Ikeda's website tells us that the work is "about the way we understand the reality of nature on an atomic scale" [51]. It presents the piece as "an immersive experience inspired by the subatomic world, which mines the notion that it is not possible to fully describe the behaviour of a single particle except in terms of probabilities". In Ikeda's words,

> The work is an orchestrated journey through sound, language, physical phenomena, mathematical concepts, human behaviour, and randomness, all simultaneously arranged and rearranged in a theatrical arc that obliterates the boundaries between music, visual arts, and performance.

The video projection is as wide as the stage, and ten video monitors are placed at the stage front. The performers appear in front of the projected image, behind the monitors. The electronic sounds are synchronised with the graphics. The visual imagery ranges from photos of electronic parts, to digitised audio waves and text. Several loudspeakers are placed around the audience.

Real-Time Effort in Sound Organisation

We can say that sound organisation does not depend on real-time effort, because the performers change their position on stage while the digital sound and image keep playing autonomously. Ikeda's audio-visual system has a large amount of autonomy, using Magnusson's terminology: in Magnusson's model the *Autonomy* axis represents the degree to which the system provides the functionality of an automata [2010].

Discontinuities

Our senses are bombarded with radical discontinuities, both visual and auditory. These are alternated with steady continuities (e.g. 00'32" - 00'35") and progressive continuities (e.g. 00'19" - 00'24"). The alternation is crucial for the sense of disruption – indeed the term radical discontinuity denotes a sudden change in the chain of stimuli.

Performer Visibility

The video projection and the images on video monitors combine as a stage scene, embracing the space around, above and bellow the two performers. Their figures are integrated as if they were part of the stage scene – all parts are equally important, the combination working as a whole.

Ecological Fit

Sound and graphics are highly synchronised, with very precise timings. There is great fit in the audio-visual mapping, but there is also a discrepancy between the mapping and the

[51] http://www.ryojiikeda.com/project/superposition

performers' physical gestures. The performers appear almost immobile behind laptops, except for occasional, slightly exaggerated physical gestures. Those gestures do not fit with the system output, and that confounds the cause-effect relationships. It remains unclear how the system works: does it depend on human agency, or is it being fed by an external flux of data?

Global Semantics

The informational dimension of the work is low because causation is concealed. The semantics of the work are more environmental than expressive because:
a) The performer's gestures are stylised, dispensing with personal expressiveness.
b) The image functions as a stage scene.
b) Sound spatialisation keeps contracting and expanding the audience's sense of space.

Performative Arena

The loudspeakers are placed around the audience, thus the performative arena is distributed. The neural impact of radical discontinuities makes us focus on our own physical body, thus the performative arena is not extended. It is not local either, because the performers are not prominent relatively to the stage scene. We see no sound sources on stage, and the system output might be independent from their physical gestures.

These parameterisations reflect how the work converges and diverges from the creative principles of this thesis. The *Ecological Fit* axes indicate a convergence with the principle of fungibility; they also show that fungibility derives from the audio-visual relationship in space, rather than from the audio-visual mapping. The axes for *Sound Organisation* and *Discontinuities* indicate that the principles of sonic complexity and visual continuity are irrelevant. Nevertheless, we cannot say that vision dominates over audition, because radical visual discontinuities are synchronised with radical sonic discontinuities, and the same for progressive continuities; the logics of the sonic construction equals the logics of the audio-visual composite.

4.4.3 Performance by Thor Magnusson

The photos and graphic in Figure 19 represent a performance by Thor Magnusson, *Fermata,* in which he uses an audio-visual system called *Threnoscope*. It was presented at ICLI 2014, lasting 30'. The *Threnoscope* system was originally conceived as a live coding piece for microtonal drone music [Magnusson 2014].

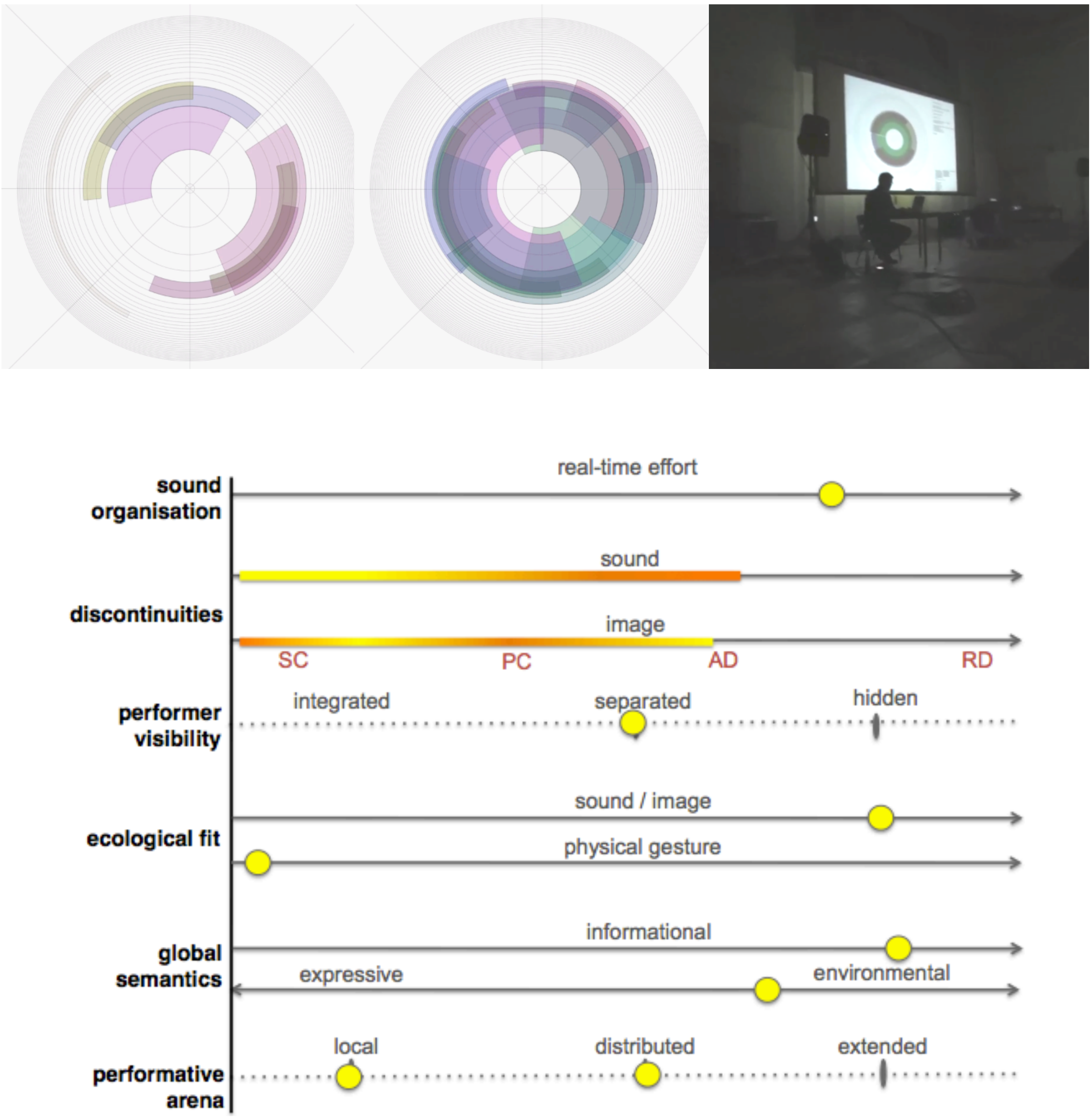


Figure 19: Thor Magnusson's "Fermata" performance, 30'
https://drive.google.com/file/d/0B9f9deWrFgQuSU1NbFJmdkZWMU0/view?usp=sharing
with his Threnoscope system https://vimeo.com/75380587

The system runs complex synthesiser instances controlled in real-time, and the graphical interface is projected on the screen. The interface is divided in two parts. One, the smallest, displays the code as it is being typed[52]. The other displays a graphical representation of the sounds in a two dimensional circular score. The score represents several sound features such as spatial location, pitch, amplitude and filtering. The circles represent harmonic series, and the lines crossing the score represent loudspeakers. A drone intersecting a

[52] The code can be edited in the text document or through a graphical, vertical time axis - the code of each drone pops up when clicked upon (see https://vimeo.com/100148092). However, Magnusson used solely the text editor in the performances I have seen.

speaker line will sound out of that speaker. In Magnusson's words, "this form of descriptive score represents the static and circular nature of the music, reducing the importance of linearity and temporality" [2014]. The drones are not ephemeral musical events, but constants that can be altered in various ways, for example by silencing, pitch shifting, filtering, and moving them around in space.

Real-Time Effort in Sound Organisation

The interaction with the system requires real-time effort and foreknowledge. Magnusson's high-level approach to live coding makes the system responsive [see Sa et al. 2015], but the timing of sounds cannot be fully controlled because real-time decisions are mediated through code typing. Hence, the graphic indicates medium-high effort.

Discontinuities

Microtonal drone music exhibits progressive and ambivalent discontinuities; its overall continuity drives us to focus on embedded harmonics, rhythms and textures. The visual projection in this work is even more continuous. There are two types of visible changes: a) the code, and b) the representation of sonic features, visualised as flat colour segments in the steady score structure. As a whole, the image prompts Gestalts of invariance and similarity.

Performer Visibility

Magnusson is visible on stage, sitting behind his laptop, separated from the visual projection.

Ecological Fit

Magnusson's code assures that the sonic changes are synchronised with the corresponding visual changes. However, the ecological fit is never absolutely conclusive for several reasons:

a) The digital cause-effect relationships are exposed with a graphic notation system and real-time code, but their understanding requires foreknowledge. Even audients who know the system and the programming language won't fully understand the cause-effect relationships, because the visible code is relatively high-level, and the system is complex.

b) The graphical score shows a clear succession of operations, and that does not coincide with how perception streams and segregates the sounds. We can see this in the video at https://vimeo.com/75380587. For example at 0'32" and 1'2" the visual change is much clearer than the sonic change.

c) There is complexity both at sonic level and symbolic level, and no abrupt audio-visual changes. We have seen that a visual target only pops out if the sonic and visual changes are both synchronised and abrupt [Van der Burg et al. 2010].
d) The performer's discreet physical gestures while sitting behind the laptop do not provide specific cues about causation.

Global Semantics

The symbolic relation of sound and image represents how the system works. This makes the informational dimension of the piece quite strong, although the individual cause-effect relationships remain inconclusive. The work also has an expressive dimension, as live coding draws attention to the performer's skills. However, the environmental dimension is much stronger, due to the sonic continuity of the drones and the dynamic sound spatialisation. We can say that the environmental dimension of the work reinforces its informational dimension, which in turn enables its expressive dimension.

Performative Arena

The performative arena is distributed, because the work is made for multichannel surround systems, implying multiple sound sources distributed around the audience. The performative arena is also local because we can see the performer on stage, and we are driven to focus on his coding skills.

In the parameterisation of this work, the axes *Sound Organisation* and *Discontinuities* indicate a convergence with the principles of sonic complexity and visual continuity. The axes for *Ecological Fit* reflect that concepts of causation remain inconclusive. Hence, we can say that the work converges with the fungibility principle as well. Given this triple convergence, vision does not dominate over audition.

4.5 Chapter Discussion

The chapter introduced a parametric visualisation model that can be used to analyse any audio-visual performance with respect to sonic expression, sensory dominance and spatial presence. Sonic expression can be inferred from three parameters: *Real-Time Effort, Sonic Discontinuities and Semantics*. Sensory dominance can be inferred from two parameters: *Visual Discontinuities* and *Ecological Fit*. And spatial presence can be inferred from how the *Performative Arena* parameter relates to all other parameters – it can be inferred from the model as a whole. The model is time-based because attention changes over time, and the axes for *Sonic* and *Visual Discontinuities* summarise how the sonic and visual changes are apprehended within a particular time-period.

The model draws from the scientific perceptual approach exposed in previous chapters, as well as from the models of Birnbaum, Magnusson and Ciciliani, which are relevant to sonic expression and spatial presence. The model of this thesis is unique in visualising sensory dominance, and the three creative principles described in chapter 3: sonic complexity, visual continuity and fungible audio-visual relationships. The model proved useful in the analysis of how different works converge and diverge with those principles.

The principle of sonic complexity dictates thresholds between the performer's control and the instrument's unpredictability, so that real-time effort conveys sonic expression. The axis *Real-Time Effort in Sound Organisation* suffices to show whether the work under discussion converges with the principle or not. The chapter exposed two ways of exploring the principle, which are challenging in terms of cognitive processing. One (Vasulka's) involves coordinating decisions about the sonic and the visual outputs, while looking at the image. Another (Magnusson's) involves mediating the interaction through code typing. Human-computer interaction is then less mediated than it would be with any acoustic input. But whereas the immediacy of the acoustic sound permits control over timing, code typing brings constraints in that aspect.

The principle of fungibility dictates a medium level of audio-visual fit, so that one senses causation without understanding the cause-effect relationships. The two axes for *Ecological Fit* show whether the work converges with the principle or not. The fungible audio-visual mapping includes synchronised and unrelated components, with enough complexity to confound the cause-effect relationships (see study from chapter 3.6). Whilst the study from chapter 3.6 looked at how synchrony conveys a sense of causation, the present investigation puts in evidence that other factors can equally convey that sensation; an example is when we perceive that visual changes are always preceded by sonic events (as happens in Vasulka's performance). Furthermore, now we know that a fungible audio-visual relationship does not require a fungible mapping. Causation can also be confounded with constantly synchronised audio-visual mappings, if the performers' physical gestures seem unrelated (as in Ikeda's performance), or if understanding cause-effect relationships requires understanding complex symbolic systems (as in Magnusson's performance).

The principle of visual continuity dictates that the visual dynamics must not exhibit radical visual discontinuities. The axis *Visual Discontinuities* shows whether the work converges with the principle or not. The image can comply with the principle by using steady continuities, or exploring progressive continuities and ambivalent discontinuities, where Gestaltist principles facilitate perceptual simplification. This can convey the environmental role of the image. The investigation of this chapter showed that the principle can also be explored so as to convey the informational dimension of a work (as in Magnusson's performance).

Evidently, my three creative principles can be explored in many different ways. The model's level of abstraction is adequate not solely to represent any audio-visual performance, but also to frame the development of an audio-visual performance language; my practical work will be described in the following chapter.

CHAPTER 5

Developing an Audio-Visual Instrument and Performance Language

5.1 Chapter Introduction

This chapter will describe how the model presented in the previous chapter informs the development of an audio-visual instrument and the physical performance setup. The model is useful to situate a number of preliminary creative decisions derived from my previous artistic path (see chapter 1.1). Regarding the principle of sonic complexity, the interaction with the instrument should require certain effort, but I prefer it to be less demanding in terms of cognitive processing than Vasulka's instrument and Magnusson's. And the sonic constructions should explore the whole range of continuities and discontinuities, intertwining expressive, environmental, and informational semantics. Regarding the principle of visual continuity, I want to maximise discontinuities at a level of detail so that the image appears organic. The image should work like a stage scene that reacts to the music. It should convey environmental semantics, as it does in Ikeda's performances. But differently, it should also extend the performative arena beyond the projection screen, expanding one's sense of space. Furthermore, I will obviously explore the principle of audio-visual fungibility, but in a different way than Ikeda or Magnusson do. I want to explore the fungible mapping, and simultaneously, a fungible audio-visual relationship in space.

My audio-visual instrument combines a zither with custom tuning and 3D software, where sound and image affect each other reciprocally. The software implements a video game engine, which is appropriate for rendering a digital 3D world and making the position of elements in that world have an impact on the spatialisation of related sounds.

Section 5.2 describes the first version of the instrument, which includes a 3D software called AG#1. The section describes how the instrument and the physical performance setup explore the principles of sonic complexity, visual continuity and audio-visual fungibility - using my taxonomy of continuities/ discontinuities and the study on audio-visual mapping/ perception (chapter 3.4). Furthermore, the section analyses the perceptual integration of sound and image.

Section 5.3 focuses on sound organisation and the principle of sonic complexity. It describes how the work evolves with a new audio software called Arpeggio-Detuning, detailing the audio analysis process, the custom zither tuning, and how the software operates based on the zither input. The section describes how the interfaces provide control and instability. It describes the adopted combinations of zither techniques and digital sounds, elaborating on how they might inform the semantic aspects of the music. And finally, the section uses my taxonomy of continuities/ discontinuities to describe how sound materials and digital constraints create a set of versatile musical forms.

Section 5.4 describes the architecture of new audio-visual software, AG#2. AG#2 implements the Arpeggio-Detuning software, extending its creative strategies to the audio-visual domain.

Section 5.5 describes how the digital 3D world and the visual dynamics explore the principle of visual continuity.

Section 5.6 describes how the audio-visual mapping explores the principle of audio-visual fungibility.

Section 5.7 describes how the digital 3D sound spatialisation explores perceptual cues, while considering the principles of sonic complexity and audio-visual fungibility.

Section 5.8 describes how sound quality is improved so as to stress the distinguishing characteristics of the instrument, and refine the expressive and environmental semantics of the work. The section describes a few components beyond the zither and the AG#2 software, and it reports an investigation about the quality of loudspeakers and amplifiers, intended to enhance the acoustic and the digital sound outputs.

5.2 Instrument Version #1

AG#1, the first version of the software from my instrument, is a modification of software created by John Klima (ironically titled *Portuguese Guitar Hero*, after the popular Guitar Hero video game series). As my modifications circumvented the original functions and aesthetics - the design in general - the software became part of an instrument, operating based on a specific, custom zither input (Photo 1). The audience does not interact with the instrument. There are no video game player paradigms or allusive icons. The performer does not face the screen. The interaction with the instrument is not clearly perceivable. There are thresholds between the performer’s control and the instrument’s unpredictability, so as to convey musical expression. And, the visual dynamics and the audio-visual relationship are architected to emphasise the sonic experience.

The parametric model of the thesis can draw general guidelines for creative strategies, but my actual compositional strategies could not have been fully planned in advance; they were rather discovered and reiterated throughout a long period of studio time. The same applies to the subsequent versions of my instrument, which draw from these strategies and take them much further.

Photo 1: Performance with the 1st version of the audio-visual instrument, St. James' Hatcham Church, London, 2012.

5.2.1 Modifying Video Game Software in Combination With a Zither

Repurposing video game technologies in combination with the zither led to exploring disparities between the acoustic sound, which is produced in immediate ways, and the digital outputs, which are mediated through software. As Magnusson notes, the body and the sound engine of the acoustic instrument are physical [2009]. Conversely, the body of the digital instrument is intrinsically theoretical - there is no natural mapping between physical input and digital output. Furthermore,

> It is the designer who decides with clear rational arguments what is revealed and what is concealed in the use of the system (…) This activity of blackboxing, of creating abstractions of activities where bodily movements and thoughts are represented as discrete chunks in time, grounds the complexity and the non-transparency of digital tools. [Magnusson 2009:174]

The coder frames and freezes his or her affordances through symbols - code. That entails a political dimension. If the "user" explores ways of unfreezing and re-framing some of those affordances, that entails a political dimension as well (see chapters 1.2.3 and 2.2.2). Particularly interesting is if the original coder actually collaborates by revealing what was concealed – then the process involves negotiation. Such is the present case: *Portuguese*

Guitar Hero was modified into AG#1 with the agreement of John Klima. I could modify the software because Klima revealed how to do it in his code (Typeless C). I changed the sonic and visual responses to the acoustic input, the digital audio recordings, and the logic of their mappings to the acoustic input; I also created a new 3D-world, including meshes, particles, "skybox", and images applied to all these elements**.**

Technically, the software developed in this research would operate upon any audio input detection. But the design specifications, digital sounds, mappings, parameterisations and structural sections are made for a specific zither, a personal tuning system, and personal zither playing techniques.

Figure 20 shows the technical architecture of the audio-visual instrument. The AG#1 software combines the video game engine *Torque*[53] and the audio library *FMOD*[54]. It detects amplitude and frequency from the zither input (photo 2, see ahead), and it renders 3D sound and 3D image accordingly. The zither playing drops visible sound emitters in the digital 3D world. These emerge as light particles, which emit pre-recorded sounds (Figure 21, ahead). The input also affects how the virtual 3D camera moves; it produces an undulating view over an unchangeable landscape. As the camera moves and the sound-emitting light particles remain where they were dropped, the digital sounds are spatialiased accordingly.

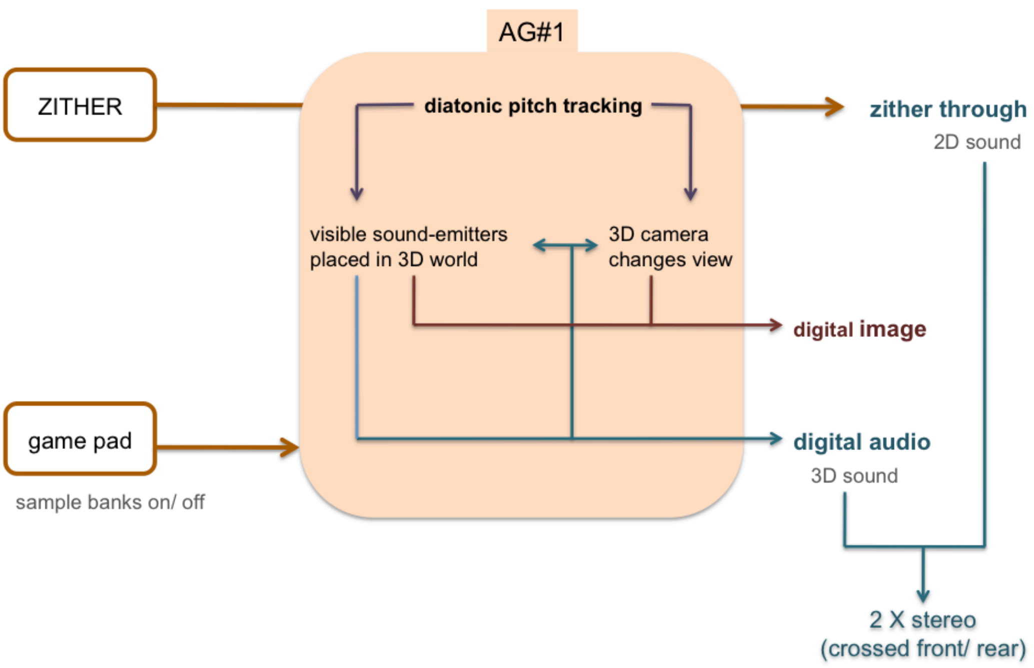


Figure 20: Technical diagram of the 1st version of the audio-visual instrument developed in this research

[53] http://www.garagegames.com/products/torque-3d
[54] www.fmod.org/

Photo 2: Zither and piezo microphone attached to its body

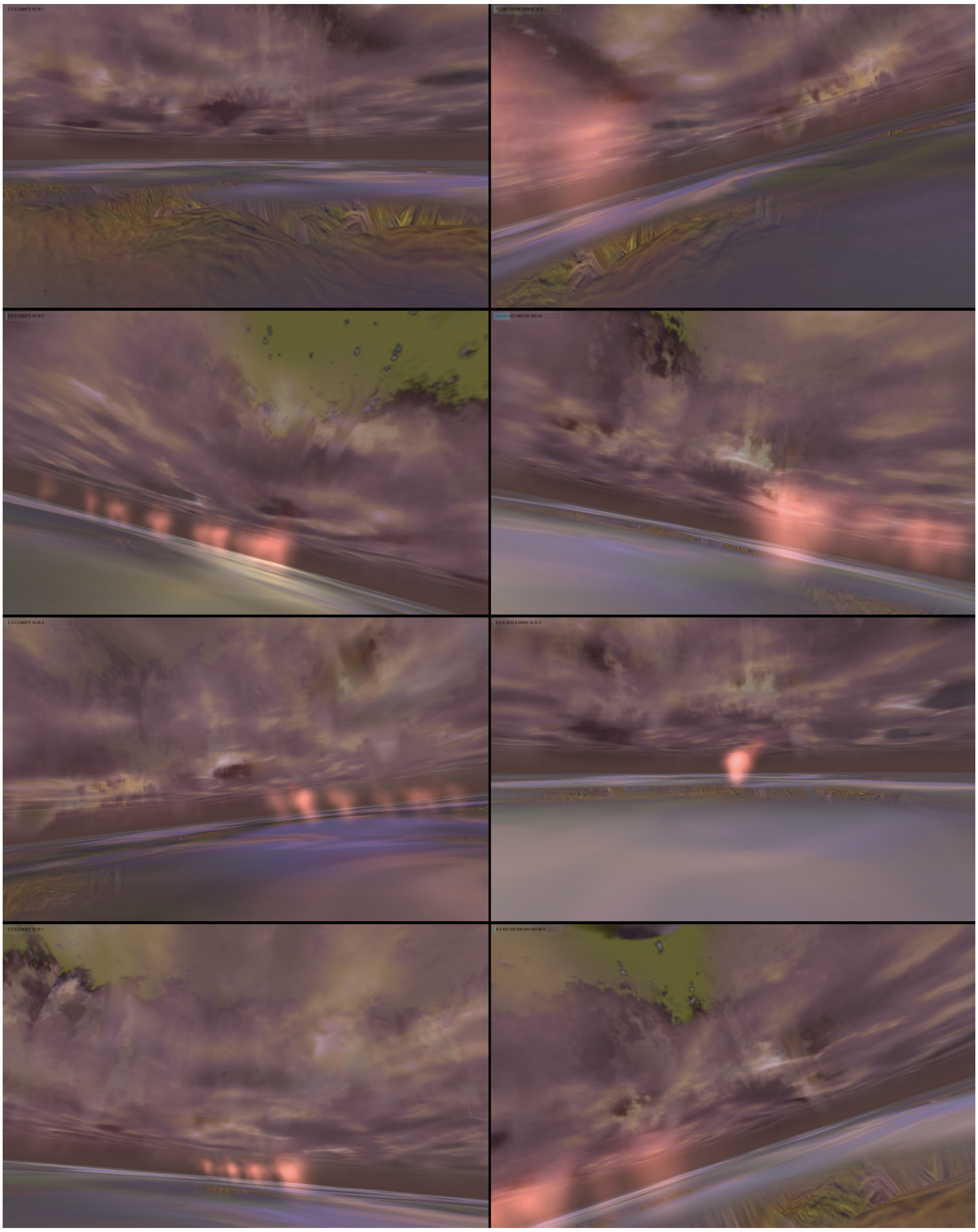


Figure 21: Screenshots from AG#1 (2012)

As a rule-based software, AG#1 embeds a set of "if-then" statements using a set of assertions, to which rules on how to act upon those assertions are assigned. This is common to Klima's *Portuguese Guitar Hero*. The difference is that the “if-then” statements, i.e. the cause and effect relationships, were then clearly perceivable so as to provide the player with control over the output. Conversely, in AG#1 the “if-then” statements are not obvious. They rule out undesired outcomes, yet their combination with each other and the acoustic, audible input creates certain unpredictability, conveying the principle of sonic complexity (Figure 22). Furthermore, this inconsistent relation between the acoustic and the digital sounds throttles the audio-visual fit. It helps to confound the cause and effect relationships, conveying the principle of audio-visual fungibility (ibid.).

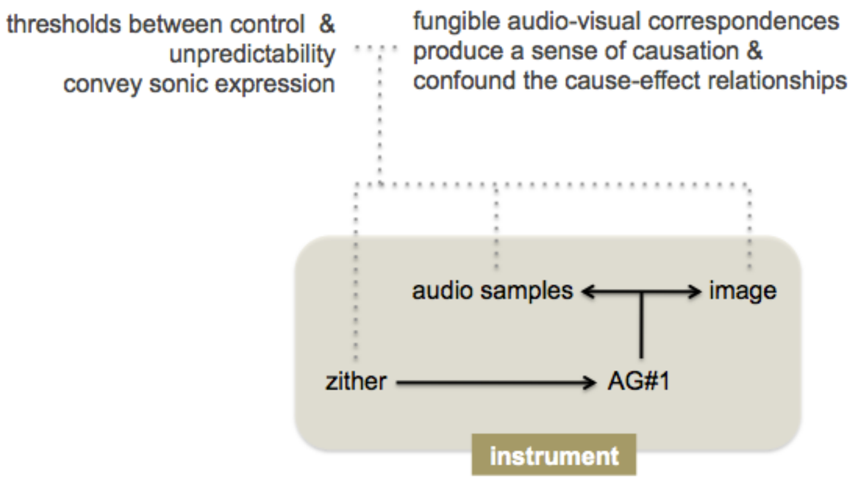


Figure 22: The disparities between the direct output sound of the zither and the digitally mediated outputs from AG#1 convey the principles of sonic complexity and audio-visual fungibility

5.2.2 Sonic Complexity and Expression

AG#1 operates through diatonic pitch tracking. The sound spectrum is divided in twelve groups of frequencies, and every detected pitch is mapped to one of these groups. Each group of frequencies is further mapped to four pre-recorded sounds, i.e. digital audio samples stored in different banks. Each bank can be activated independently, through game pad buttons. The relation between acoustic and digital sound may at times be opaque to the audience, and nevertheless transparent to myself as a performer. At other times, it becomes opaque to me as well.

AG#1 operates according to mappings, but in performance the outcome is not fully predictable. Firstly, the actual resilience of the zither playing exceeds its codified terms – the digital processing does not handle *all* the acoustic information, since it is based on sampling the input. Secondly, the software at times considers elements that one is not perceptually aware of, and it responds accordingly. An instrument combining acoustic and digital sound can explore these disparities so as to create performative tension; as the instrument thresholds control and unpredictability, the interaction is volatile, which potentiates expression.

Several technical aspects convey performative control: the zither playing is audible regardless of whether the software detects it or not; the audio sample banks (not the digital audio recordings themselves) are activated or muted through the game pad; and clear notes generate predictable response.

Other technical aspects convey performative instability. The digital sound may overlap with its acoustic cause, or appear immediately after, or overlap with subsequent sounds. Whereas playing *mezzo forte* or *forte* activates digital audio samples, playing *piano* does not – and while fluidly playing the zither, I do not fully predict where that threshold is. When the acoustic sound has no particular tonality – e.g. when zither strings are plucked, picked or strummed - the software response is unpredictable. The zither tuning is somewhat ambivalent and a few of its strings are purposefully aged. Thus I do not fully predict which audio sample will be activated, and a single string may activate several digital sounds mapped to different pitches.

Each digital audio bank gathers a particular type of pre-recorded sounds, and each bank is intended to combine with a particular type of zither techniques. Each combination can be thought of as a musical form, corresponding to a specific range of dynamic and semantic possibilities. Those possibilities are not really limited, in spite of the digital constraints inherent to any mapping. For example, I can overlap the audio banks and/ or change zither playing techniques at any time. The point is to have four types of musical vocabulary, each conveying a different type of sonic construction.

It is useful to recall how the model of this thesis parameterises semantics (see chapter 4.3.4). In using the term *Expressive Semantics*, it narrows Pressing's notion of "expressive sounds", i.e. all kinds of music and song [1997]. In our model, the term relates to Emmerson's "local functions" [2007] and Ciciliani's "centripetal" performance tendencies [2014]: it denotes a focus upon the performer. The term *Environmental Semantics* relates to Pressing's "environmental sounds", e.g. animal calls, wind sounds and the noises of machinery [1997]. It also relates to Emmerson's "field functions" [2007] and Ciciliani's "centrifugal" performance tendencies [2014], which imply that the environment becomes more important than the performer. Finally, the term *Informational Semantics* relates to Pressing's "informational sounds", e.g. speech, alarms and sonified data [1997]. In the context of music, the term embraces all situations where the sonic construction evokes something beyond itself. This means that the informational dimension of a sonic construction can inform its expressive and environmental dimensions. For example, a recording of singing birds evokes birds, and a piano recording leads us to imagine a piano and a pianist although we are solely hearing the sound.

Returning to my work, the sounds stored in bank #1 are long, with digital timbres. With this bank I dribble the zither (generally, dribbling means to flow or allow to flow in a thin stream or drops), or play sparse notes that function as punctuations in the dense soundscape (dribbling also means trickling, i.e. to flow in small streams). While the digital sounds create sonic density, the zither sounds merge and submerge in that mass of sound: the focus shifts away from the performer, so as to create environmental semantics.

The sounds from banks #2 and #3 are short recordings from the zither, in which I pluck, pick and strum the strings. I activate these banks when I am using the same techniques in real-time, so that the acoustic and digital sounds are at times undistinguishable. The sonic changes are much faster than with bank#1, creating more unpredictability; this draws attention to the performer's skills, conveying expressive semantics. At times one recognises the zither timbre, and other times not; this creates a variable level of informational semantics.

The sounds from bank #4 are longer recordings from the zither, played with bow. I activate this bank when I use the bow in real-time as well. The acoustic input is indistinguishable from the pre-recorded sounds, and the combination sounds like a string orchestra. As one recalls well-known instruments like violin or cello, the music has an informational semantic dimension. As the seemingly recognisable timbres evoke skills for playing those instruments, the music has an expressive semantic dimension. And because one feels surrounded by many musicians playing simultaneously, the music also has an environmental dimension.

The general sequencing of the musical structure is sketched with a graphic score, but real-time decisions and timings are primarily visceral. Graphic scores are not a matter of research in this thesis, but it is worth showing my personal notation system because in a way, the instrument adopts analogous compositional strategies. Figure 23 shows a score from a performance with this first version of the audio-visual instrument. My notation system, which I use for more than twenty years, consists of a drawing and vertical lines that map its different parts to musical sections, playing techniques and instrumental combinations over time. The drawing suggests densities, textures, and transitions between sections, leaving clock-time and musical phrasing open, as well as many other aspects of the sonic construction.

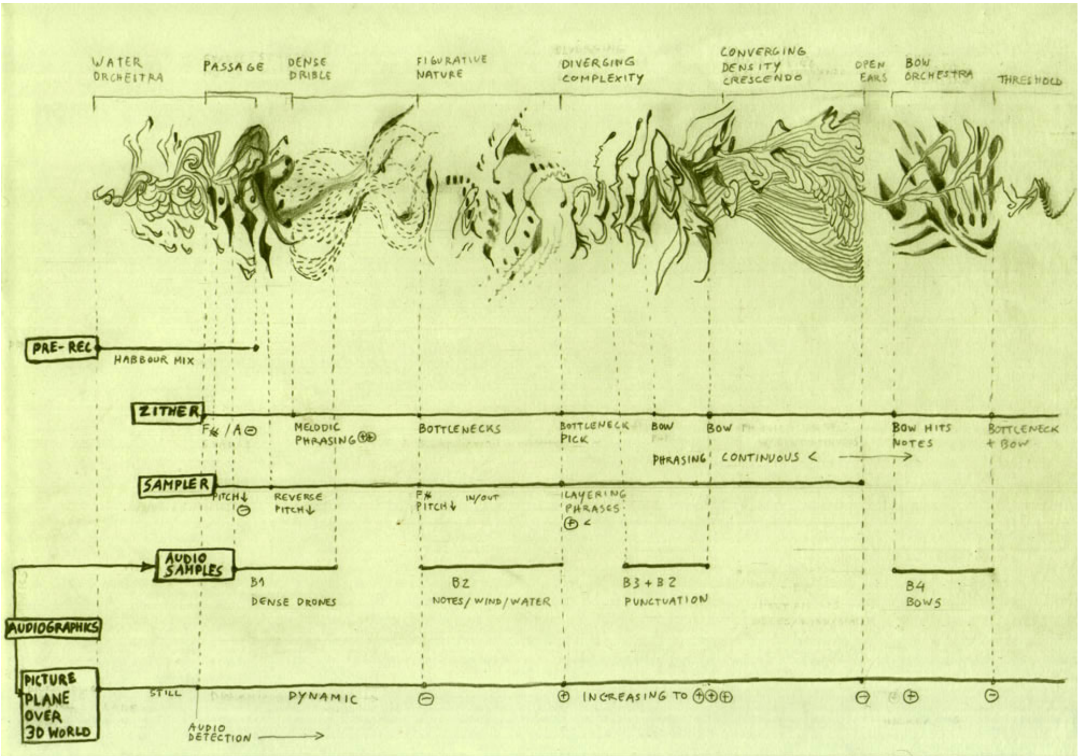


Figure 23: Graphic score for a performance with the audio-visual instrument (2011)

5.2.3 Visual Continuity

The projected image represents an imaginary deserted landscape, hence its semantics are predominantly environmental. In turn, the environmental semantics derive from informational load: the image represents something beyond itself, namely a 3D world where things occur.

The instrument addresses the principle of visual continuity because the existing visual discontinuities do not automatically attract attention. Yet, the image is not really continuous. The camera view over the 3D world changes, light particles appear and faint, either isolated or drawing paths through the world. Such variations subordinate to an overall continuity, as a few Gestaltist principles facilitate visual simplification.

There are no changes of scenario in the digital 3D world, and the light particles placed there are monochromatic and similarly shaped. This conveys the Gestalt of invariance (an object is recognised independent of rotation, translation, scale, elastic deformations, different lighting, and different component features) and the Gestalt of similarity (similar elements are grouped as part of the same form). Accumulating particles create visual paths, with subtle shifts in the foreseeable logic of events. These shifts create ambivalent discontinuities, which at low resolution become continuous.

The functionalities of the 3D camera are adapted from a first-person hovercraft racing video game. Using the car driving metaphor, the “steering force” is low, there is no “break” and the “inertia movement” is large. This means that the 3D camera curves smoothly, creating an undulating view over the 3D world. Pitch is mapped to four camera target directions (up, down, left, right). In-between two audio input detections, the camera moves at consistent speed, progressing towards a point of limit (progressive discontinuity). That point of limit coincides with the next audio detection, which makes the camera curve smoothly in another direction.

5.2.4 Fungible Audio-Visual Mapping

As the audience looks at the image, perception binds sonic and visual shapes that change or move simultaneously (Gestalt of common fate), as well as those that change adequately proximal in time (Gestalt of proximity/ sequential integration). Yet, one remains unsure about the base cause and effect relationships, as is characteristic of the fungible mapping. The mapping combines synchronised and apparently non-related components, exhibiting inconsistency and complexity so as to confound the cause and effect relationships. The combination invites perception to form inconclusive audiovisual objects.

A few synchronised audio-visual components produce causal percepts. The light particles and 3D camera direction shifts occur immediately upon audio detection. In addition, the speed of sound spatialisation reflects the speed of the 3D camera, creating a seemingly natural relationship (Gestalt of common fate).

Other factors bring complexity and inconsistency, confounding the actual cause and effect relationships. Whereas the zither is always audible, the software only responds above certain amplitude. The dynamics of the sky and the water from the 3D world are not mapped to sound. The biggest visual changes (emerging light particles and prominent 3D camera direction shifts) are consistent with zither input detection, but the digital sounds are inconsistent with that same detection. I can stop their activation, interrupting the cause-effect relationship. And several audio recordings start with random delay. Moreover, using digital recordings of the zither makes the acoustic and the digital sounds at times undistinguishable.

5.2.5 Performative Arena & the Fungible Audio-Visual Relationship in Space

The semantics of the work extend the performative arena beyond the physical performance room, but the action concentrates in the physical space. AG#1 explores perceptual cues so as to convey the audience's spatial presence in the digital 3D world[55]: the perspective view from the virtual camera creates a sense of visual depth, and the visual components synchronised with audio detection produce causal percepts. I stand in front of the screen, in the threshold between the physical and the digital 3D world. The performative arena embraces both spaces, yet in the digital the events occur at slower pace than in the physical. The moving image is projected over me, and I dress white (photo 3). This reinforces its environmental role, as a stage scene that morphs with the music.

Photo 3: Performance with the 1st version of the audio-visual instrument at In-Sonora, Madrid, 2011

[55] See chapters 2.2.3 and 2.6, which describe how 3D environments and film explore the perceptual mechanisms through which we naturally experience the world, so as to convey a sense of spatial presence beyond the screen.

Software configuration requires interfacing the instrument through the screen, but in performance my interaction is solely through the zither and the game pad. I draw upon sonic constructions, sensing the image solely as light. Since vision has an impact over audition, the audience's experience differs from mine. Nevertheless, the disparities are limited because the visual motion subordinates to the sonic motion.

As we consider the audio-visual relationship in performance, sound diffusion is of great importance. Usually, 3D sound is used with multi-channel systems so as to create unequivocal links between sound and image. The sound is rooted dynamically to a set of loudspeakers, so that the spatialisation of sounds corresponds directly to the related visual elements on the screen. This creates a sonic simulation of the digital 3D space, based on a seemingly natural relationship between sound and image. Maximising the audio-visual fit in this way would counteract the principle of fungible audio-visual relationships.

My work dispenses with multichannel diffusion systems. The software outputs 3D sound into two stereo systems, placed around the audience, crossed rear to front. As the digital sounds move through the loudspeakers, their speed corresponds to the speed at which the camera moves relatively to the light particles on the screen, i.e. the virtual sound emitters in the digital world. This speed similarity creates a level of audio-visual fit, which conveys perceptual binding. Yet there is no audio-visual correspondence in terms of directionality, because one stereo system diffuses the audio signal inversely to the other. Thus, the cause-effect relationships can be confounded.

Physical gesture is equally important to the audio-visual relationship. A clearly perceivable interaction with the instrument would counteract the principle of fungibility. Quoting Simon Emmerson, "we may opt for a more ambiguous relationship, mixing some directly perceivable cause-effect chains with a) relationships between performer gesture and result which the performer may understand but the audience not, or b) relationships of a more 'experimental' nature the outcomes of which may not be fully predictable" [2007:91].

5.2.6 Perceptual Integration of the Sonic and the Visual Motion

The sonic motion oscillates between points of highest intensity (radical discontinuity) and lowest intensity (steady continuity). The visual motion is much narrower - it produces an overall sense of continuity, with progressive and ambivalent discontinuities.

The visual motion never reaches high intensity: monochromatic light particles and smooth camera shifts are the biggest visual changes. Those changes do not skew surreptitious sounds, inclusively because they are synchronised with amplitude detection, and the high detection threshold assures that the corresponding zither sound is relatively loud.

Visual dominance is avoided because there are no disruptive visual changes, and the audio-visual relationship does not produce conclusive concepts of causation. As the audience looks at the screen, deliberate attention can increase the intensity of visual details. But the ventriloquist effect does not occur (perceptual dislocation of the sound source toward the visual target), because it solely happens with automatic attention [Bertelson et al. 2000]. The spatialisation of sounds puts the focus in the physical space, rather than in the digital space beyond the screen.

As the performative arena links the physical and the digital space, the image creates an organic, reactive stage scene. Perception is driven by sound organisation, in its fluid oscillation between tension and release. The stimuli panorama leaves attention predominantly under individual control, except at points of radical sonic discontinuity. Kubovy and Yu argued that visual multistabile stimuli do not affect how we perceive auditory multistabile stimuli, nor vice-versa [2012]. Yet, as one focuses on the complex relations between the sounds, that causes changes in perceptual resolution. Attention causes working memory to optimise resolution according to its target [Knudsen 2007]; perceptual resolution determines the scale of each representation, and consequently, the ratio of perceived continuities/ discontinuities. Since audio-visual percepts are synthesised at that changing resolution, the complexity of sound organisation also affects vision.

5.2.7 Notes: Towards an Expansion of Creative Strategies

So far, the chapter described the first version of my audio-visual instrument, as well as the physical performance setup in space. This early stage of practical work corresponds to the emergence of a personal audio-visual performance language, which explores the creative principles of sonic complexity, visual continuity, and fungible audio-visual relationships. The work explores functionalities from 3D software meant to create video games, while taking advantage of disparities between acoustic sound, digital sound and digital image.

The work should be enhanced with new musical forms and compositional strategies, as well as improvements regarding sound amplification and sound diffusion. Gathering pre-recorded sounds in different audio sample banks is useful for the dramaturgical structure of sonic constructions, because one bank can convey *Expressive* semantics and another *Environmental* semantics. Furthermore, the image should become more organic at a level of detail, and that should reinforce the environmental rather than the informational semantics of the work. And regarding the audio-visual relationship, the work should be enhanced with a new, more organic fungible mapping, as well as particular 3D sound spatialisation strategies for inverted stereo diffusion.

There are certain challenges inherent to my three creative principles. Firstly, I must be careful with increasing sonic complexity; this first version of the instrument already explores the fringes of performatory control. Secondly, the image is desired to create a dynamic environment, but visual changes should not become too discontinuous or polyphonic. And thirdly, the fungible audio-visual mapping can be more complex, yet it requires a careful balance between synchrony, inconsistency and complexity.

5.3 Sound Organisation: Arpeggio-Detuning

This section describes practical work focused on sound organisation, namely my instrument version 2.1. I will describe an expansion of creative strategies with new, rule-based audio software, which was later also adapted to audio-visual software. Along with this description, I will make observations relevant to the parameterisation of interaction, sonic dynamics and semantics.

The new audio software is called Arpeggio-Detuning, as it applies arpeggio and pitch down algorithms to pre-recorded sounds. Similarly to AG#1 (the audio-visual software described in the previous section) the Arpeggio-Detuning operates based on amplitude and pitch analysis from the acoustic zither input. But very differently, the Arpeggio-Detuning considers the principle of sonic complexity at a low level in the digital architecture, rather than solely at a high level in parameterisation.

The zither was originally made for the Western 12 notes scale. My customisations and tuning convey my visceral relation with pitch, which makes tones, microtones and all chromatic shifts in-between equally important. These customisations were refined along with my zither techniques, which are not traditional at all. I wanted the new software to create chromatic shifts based on the zither input, so as to develop a musical idiom where acoustic and digital sounds at times converge in unison, and other times unfold in independent directions.

John Klima implemented my specifications with C++, combined with a cross-platform, mobile-based hardware abstraction layer called *Marmalade*[56], which is used in video games; and an audio library called *Maximilian*[57]. I calibrate digital configurations and sounds on a PC. In performance, my interfaces are the analysis from the zither input and an iPad touch screen.

Photos 4 and 5: Setup including zither, piezo microphone, compression pedal & iPad with the Arpeggio-Detuning software

5.3.1 Amplitude and Pitch Analysis

Amplitude and pitch analysis require collecting and converting sound signal into data. Amplitude, or loudness, is the maximum displacement of the sound wave from a zero level position; it is unrelated to frequency, which is inversely proportional to wavelength. The amplitude of a digitised sound wave varies between -1 and 1, being 0 equal to silence, but algorithms for amplitude analysis usually consider the absolute value so that amplitude varies between 0 and 1 (Figure 24). The software collects a number of samples from the audio input (discrete analysis) and calculates the average amplitude value. The number of collected samples corresponds to the size of the sample buffer. In the Arpeggio-Detuning software, the sample buffer for amplitude analysis is small, thus the average value is accurate. The software can respond inconsistently to amplitude variations, but that is because the amplitude detection threshold is high; it is a constant, which I set prior to performance.

[56] iOs/Android development kit, https://www.madewithmarmalade.com/
[57] By Mick Gierson, https://github.com/micknoise/Maximilian

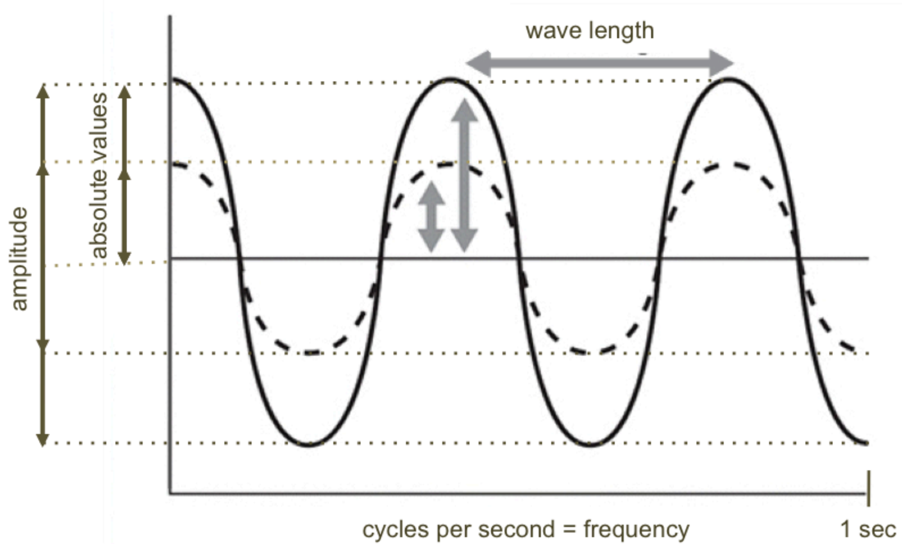


Figure 24: Two sine waves with different amplitude, but same frequency and wavelength

Pitch, i.e. frequency analysis is far more complex than amplitude analysis. Sounds include a range of frequencies (spectral shape), from which the software must extract the fundamental based on mathematical calculations. The question is what the notion of "fundamental" entails.

Usually, pitch analysis involves windowing the audio signal, as happens with amplitude analysis. A Fast Fourier Transform (FFT) function produces a series of values that represent the amount of energy, in a range of equally spaced frequency bands over a given number of samples. The frequency resolution depends on the number of frequency bands (called FFT size). There are important limits on the information we can get from the Fourier transform of a time-limited signal extract, particularly with a small FFT size. Pitch is audible frequency, and frequency is the rate per second of a vibration constituting a wave. A very short fragment of signal does simply does not contain sufficient information about periodicity. Yet, a large window creates latency, and even a little amount of latency can break the sense of immediacy in interactive systems. Musicians can detect reaction latencies of 20-30ms in musical instruments [Mäki-Patola and Hamäläinen 2004, Adelstein et al 2003], and the accepted target with interactive audio systems is latency under 10ms [Freed et al 1997].

Usually, the design of interactive digital systems endeavours to negotiate frequency resolution vs. latency so as to provide a sense of immediate interaction. Chapter 2.5.1

pointed out how designs seek to mimic how we naturally perceive the world, so as to convey embodiment - the medium becomes "transparent". Chapter 3.3 elaborated on how musical expression can benefit from the combination of embodiment, and separation between performer and instrument.

The construction of musical time requires a sense of immediate interaction, but that is not a concern in my software design; immediacy comes from the acoustic zither. Rather than attempting software to create a sense of immediacy, my creative strategies seek to emphasise, i.e. take advantage of the disparities between human perception and digital analysis.

In fact, these disparities are unavoidable: whereas software operates based on mathematical calculations, humans sample and process the information based on attention, cognitive principles, and cross-sensorial context. Our percepts are always informed by expectations and concepts derived from past events. Conversely, software isolates the momentary input, and it responds accordingly. The discrepancies are tangible. For example, a sound may vary in pitch during attack, sustain and release, and nevertheless, we form Gestalts so as to group those pitch variations and segregate the sound from the soundscape [Bregman 1990]. In contrast, the software slices the spectrum according to a buffer size, which may lead to overtones or resonance frequencies to be extracted as fundamentals. Also, an overtone can become intense as attention causes working memory to optimise perceptual resolution, without the pitch being fundamental according to mathematical formulas.

5.3.2 The Zither Tuning and the Arpeggio-Detuning

Given the discrepancies between human perception and digital analysis, one can create sonic complexity and unpredictability with purely rule-based software, operating based on an acoustic, audible input. The combination can convey a musical language formed of surreptitious chromaticisms and timings, where expression comes from avoiding easy developments.

My zither strings are from bass guitar, electric and acoustic guitar, and Portuguese guitar. Some are purposefully aged, which makes their timbre less shiny. As I started developing the Arpeggio-Detuning software I adopted a consistent, personal tuning to B 440Hz: around Bb, B, D, Eb, E, F and F#, but never exactly. The strings are played in any combination with hands, bottleneck, pick or bow. If the zither was plugged into a guitar tuner, the tuner would display a succession of different values upon a single string or chord.

The Arpeggio-Detuning software uses two streams of data from pitch analysis, as well as their mathematical difference (Figure 25). One data stream corresponds to the detected, fundamental frequency, calculated with a Fast Fourier Transform. The other corresponds to the nearest tone/ half tone - A, A#, B, C, C#, D, D#, E, F, F#, G and G#. These tones/ half tones are not played back. They are further mapped to pre-recorded sounds (octaves are disregarded), so that each audio input detection causes a corresponding pre-recorded sound to play back two times. The result is not repetitive because the second play back is pitched down, i.e. detuned according to a variable value, namely the difference between the detected pitch and the closest tone or halftone.

We can think of the Arpeggio-Detuning as a tuner: it displays the difference between the detected pitch and the closest tone/ half tone. Except, the Arpeggio-Detuning software “displays” the difference in audible, rather than graphic ways.

Detection:

The Arpeggio-Detuning collects 4096 samples (buffer size) from the zither input, at 8000 Hz. Once the buffer is full, the data is run through a “Second Order Filter” and a “Hamming Window”, or (alternatively) a “Han Window”. The data is analysed based on what we call a “full spectrum table”, or FFT table. In this table, the detected frequency spectrum (20Hz to 4000Hz[58]) is divided in 4096 equal parts (the table size equals the buffer size). The full spectrum table is defined by this function:

freqTable[i] = (SAMPLE_RATE * i) / FFT_SIZE

where "i" is the index from 0 to FFT_SIZE and SAMPLE_RATE is 8000.
Implementing the "libfft", Fast Fourier Transform library[59] the FFT provides the highest magnitude frequency from the zither input - which I have previously called the *detected* pitch.

Detected pitch to note mapping:

The detected pitch is then mapped to one of 128 frequency values, in what we can call the

[58] The Nyquist theorem states that in order to represent any sinusoid, we need to sample it twice at a minimum. Hence the Nyquist rate is half the sampling rate.
[59] Copyright (C) 1989 by Jef Poskanzer

“Note-Mapping Table”. These frequency values correspond to tones and halftones from the western note scale; the piano has 88 keys, and the table adds a few tones/half tones above and below. Each division or "note" is arrived at by this equation:

$$pitch = (440.0 / 32.0) * (2 ^ {((i-9.0)/12.0)})$$

where "i" is the "piano key" from 0 to 127.

If, say, the absolute frequency of a certain C is 694.5 Hz, the program finds the index into the FFT table closest to that frequency, and maps it to that particular C in its particular octave. From a tuning point of view, that would be the measure of precision. Naturally, the larger the FFT table, the greater the precision.

Recapitulating, the FFT narrows the acoustic input to 4096 values, and then, the mapping between the detected pitch and the closest tone/ half tone narrows the input to 128 values. Moreover, the note frequencies do not result from dividing the audible spectrum in equal parts; the division is logarithmic.

The next step in pitch analysis is called *chromagram*. It narrows the 128 frequency values of the Note-Mapping Table further, to 12 values. Performing a modulus 12 (modulus is the remainder when you divide by a number) of "i" above (e.g. i % 12) gives us the C to B standard tone/ half tone "name" within the total octaves included within the 128 notes. In other words, any detected pitch will give the “name” of the closest tone/ half tone - one of twelve. These names establish our “frequency groups”. In this way, every detected frequency is always mapped to one of twelve pre-recorded sounds (or more, if we have more than a single audio sample bank).

Detuning value

Once one of the 12 notes is detected, the software plays the corresponding audio file. Again, the file has no frequency correspondence to the note - rather, based on compositional criteria, a sample is carefully selected to correspond to the note. The sample is then played a second time, pitching the sample down by the difference in frequency of the "pure" note and the result of the FFT analysis. For example, if the FFT result for a certain detection cycle is 720 Hz, and the closest pure note is C at 694.5 Hz, the "C" sample will play a second time at 25.5 Hz lower frequency. It is a simple file playback modulation, thus the playback duration of the sample is extended as well.

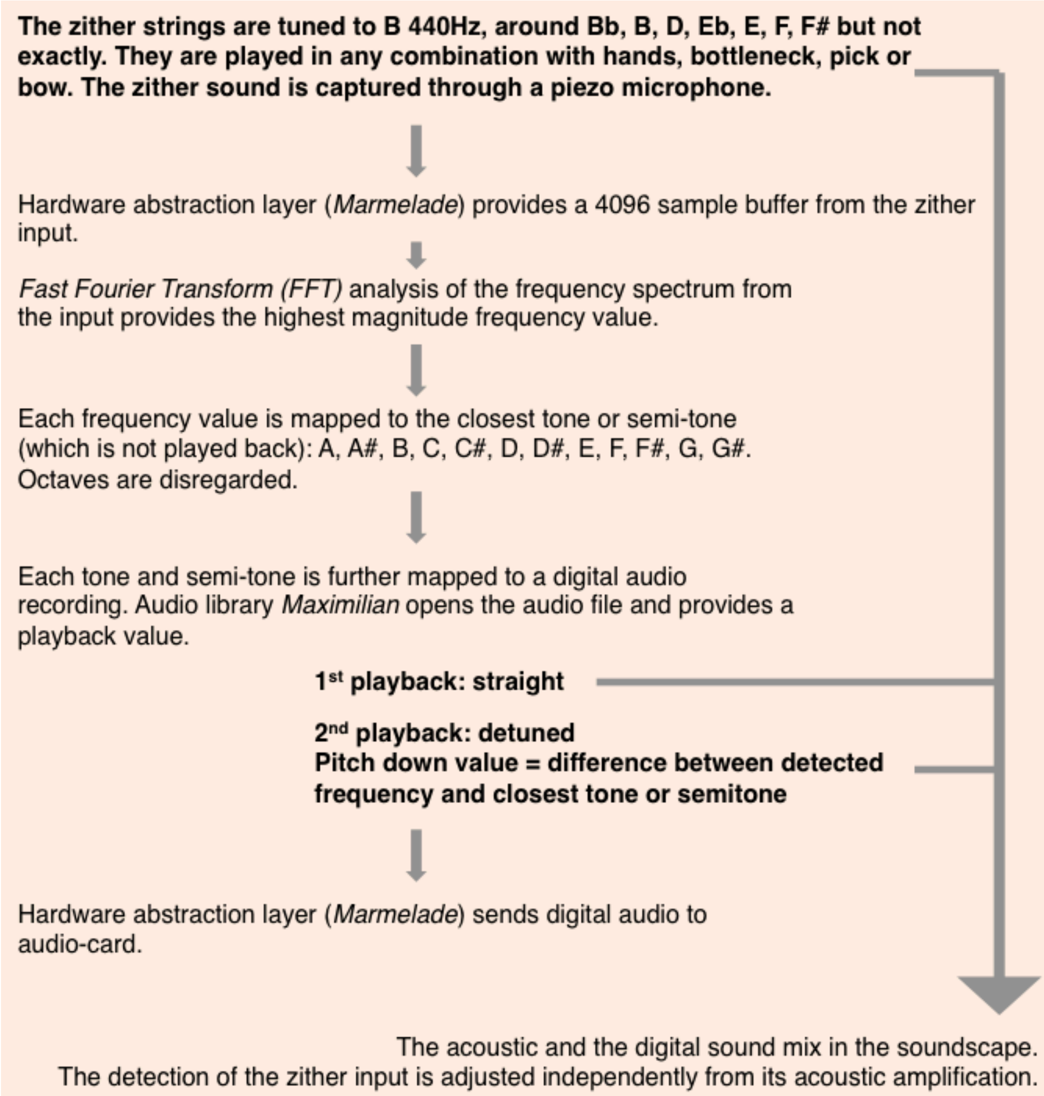


Figure 25: Technical diagram - zither with Arpeggio-Detuning software

5.3.3 The Interfaces: Technical Aspects of Control and Instability

The Arpeggio-Detuning is configured on a PC, and downloaded to an iPad (which is obviously more portable, but does not allow any file management). The digital sounds are stored in three banks, each containing twelve pre-recorded sounds. There are two interfaces for digital audio when I play the instrument: the pitch analysis from the zither input and the iPad touch-screen. Pitch analysis brings unpredictability, namely regarding which specific digital sound is played back upon detection, and how much pitch down is applied to its second playback. Yet the iPad touch screen enables me to choose musical sections: input detection will only trigger sounds from the currently activated audio sample bank.

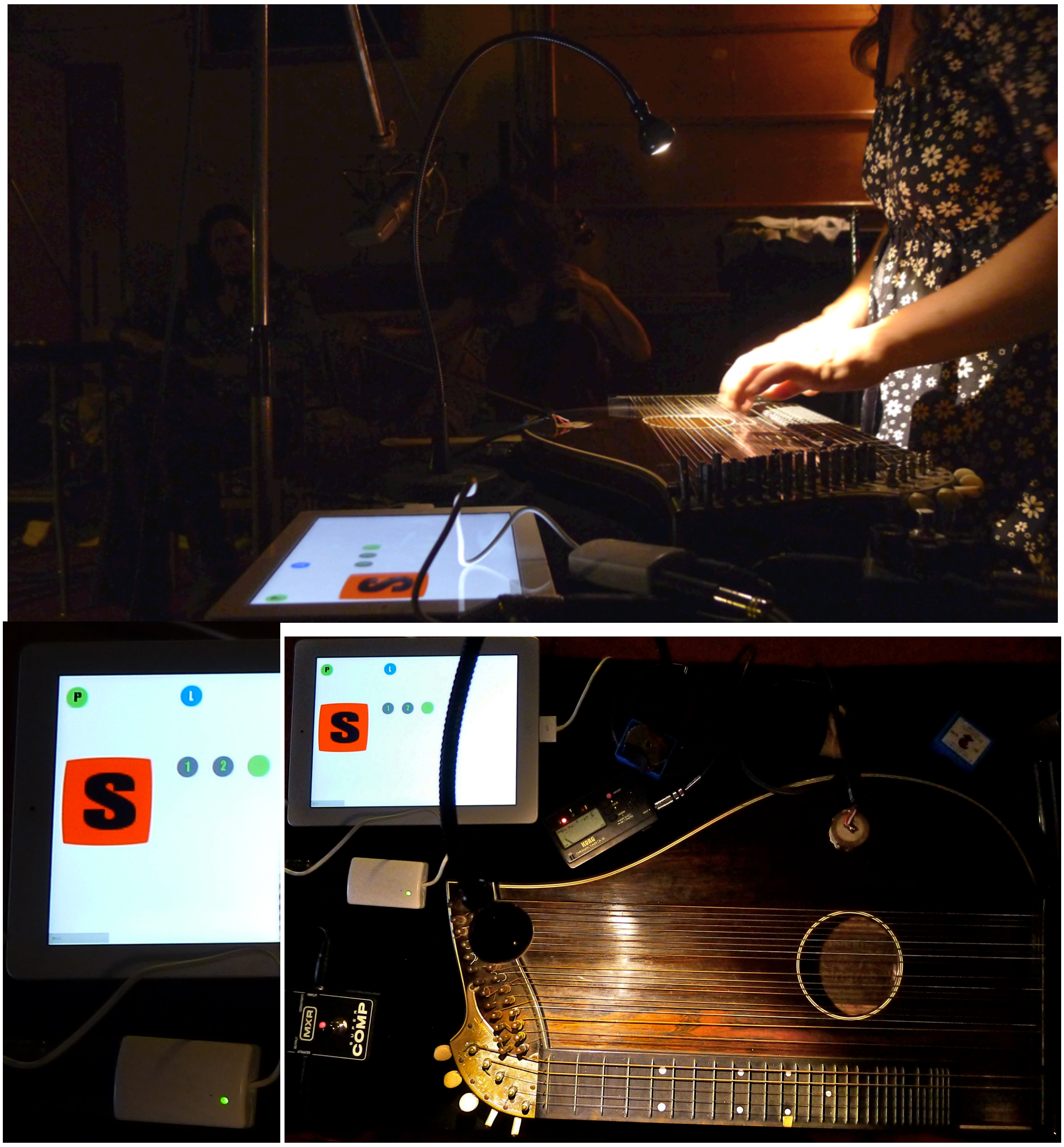


Photo 6: Playing the zither with the Arpeggio-Detuning software, 195 Lisbon Studio, 2013

The iPad screen displays seven digital buttons (photo 6), three of which are not used during performance. One button loads all three sets of digital sounds to RAM (Random Access Memory). Pre-loading all digital sounds prior to their actual activation minimises latency, i.e. the time it takes to complete an operation upon sound input. A second button on the touch screen initiates amplitude and pitch detection. A third exits the application.

The other four digital buttons are used while playing the instrument. One switches digital audio on and off. Switching it off does not interrupt the audio playback in a sudden way; the software finishes playing back the previously activated sound. Conversely, if I switch from one set of digital sounds to another, any currently playing sound is immediately interrupted. A second button activates/ deactivates sample bank 1. A third activates/ deactivates sample bank 2. And a fourth activates/ deactivates sample bank 3.

Compared with the game pad, which I used with the previous software AG#1, the iPad has the advantage of showing whether the digital audio playback is activated or not, and which audio sample bank is currently activated. The downside is that there is no tactile feedback. Digital buttons have to be touched precisely in the middle, and fingers must not slide; in short, the interaction depends solely on vision. However, in fact it does not really make a difference if I use an iPad, a game pad or a computer keyboard, because I touch the buttons sparsely - solely to initiate and terminate musical sections.

Several technical aspects convey control over the sonic construction. The zither playing is audible whether the software detects it or not. A digital constraint is implemented, which neutralises audio activation whenever the detected pitch is mapped to the same sound than the previous detected pitch. Each sample bank is activated or muted through a separate, digital switch. Another switch turns digital audio globally on or off.

Other technical aspects convey performative instability. Because the amplitude detection threshold is high, as I am fluidly playing the zither I do not fully predict where the detection threshold is. Given my personal tuning, the piezo amplification, and the process of pitch analysis, a single string can activate several audio samples, each mapped to different tones/ halftones. And when the acoustic sound has no particular tonality (e.g. when zither strings are plucked, picked or strummed) the software response is particularly unpredictable. Also, I know the source-sounds stored in audio sample banks, but they become new sounds when pitch down is applied.

5.3.4 Zither Techniques, Digital Sounds and Semantics

I developed specific zither playing techniques with each three audio sample banks, creating a set of versatile musical forms. The parametric model described in the previous chapter also led me to pin down certain aspects relevant to the semantics of any sonic construction.

Table 4 shows the adopted combinations of zither playing techniques and digital audio sample banks.

	Audio sample banks	Zither techniques
Combination #1	bass guitar, ocean waves, water drops, thunder, wind	dribbling, bow
Combination #2	dobro, bass guitar, zither (bottleneck and pick)	hands, bottlenecks, pick
Combination #3	piano notes and digital timbres	bow, bottleneck

Table 4: Combinations of sample banks and zither techniques (http://adrianasa.planetaclix.pt/research/ArpeggioDetuning.htm#music)

In these musical forms, some digital sounds are more recognisable than others. We can say that recognised sounds are *figurative*: they recall an instrument, a situation, or a place. The term figurative should here be understood as it is painting: it refers to any type of representation. But whilst the word 'representation' seems to emphasise a relation with real-world object sources, the word 'figurativism' seems to dispense with any distinction between "real-world" and "evoked world". Hence, the term is appropriate to describe the informational semantics of a sound or a sonic construction.

Many sounds from sample bank 1 are quite figurative: sounds of water and thunder recall nature. But importantly, semantics also depend on the performative context and individual association mechanisms. For example, many sounds from sample bank 2 are recorded from the zither, yet they may recall different instruments such as saxophone or guitar. They become acousmatic because I am not creating them in real-time: the audience lacks a visual reference to assess their actual source. Given this lack of visual references, even the most figurative sounds can become abstract when merged with other sounds.

My music invites perception to navigate in-between:
a) Figurative meanings, namely when the source or the cause of sounds is recognisable;
b) Ambivalent meanings, namely when the sound is associated to several possible causes;
c) Abstract meanings, namely when we strip the sound from causes or meanings.

5.3.5 Musical Forms and Digital Constraints

Given my three combinations of zither techniques and digital sounds (table 4), this subsection uses the taxonomy of continuities and discontinuities to describe how digital configurations and sonic characteristics shape musical motion. I will specify how sound materials, playing techniques and instrument configurations enable and condition the resilience of musical forms, both at a dynamic and a semantic level.

In my instrument, the digital sounds create certain unpredictability, and the acoustic immediacy of the zither enables the music to shift in good time and direction. Unpredictability must be calibrated so as to convey expression without being disruptive. Radical discontinuities attract attention automatically, creating points of high intensity in musical motion. They need to be sparse and very precise, which requires me to have direct control over the instrument; hence, I solely create radical discontinuities with the zither. The digital sounds never create radical discontinuities, because they are always preceded by a zither sound, and the amplitude detection threshold is high.

Each audio sample bank contains twelve pre-recorded sounds. A digital sound can be short or long, simple or complex. It can also have a shorter or a longer attack (i.e. the time it takes to reach the maximum amplitude), sustain, and release (i.e. decay time). Short attacks and/or releases create greater discontinuity than long ones.

I do not have direct control over which sound is triggered upon detection, which means that each audio sample must combine well together with the other eleven samples in the same bank. Each sample can become part of many different musical shapes. As the samples combine with each other and the acoustic sound, perception will stream and segregate the component parts depending on the musical motion as a whole.

The musical shapes can become quite complex, and that is likely to create density. A single musical gesture on the zither can activate several digital sounds, and each is played back twice. Moreover, the pitch shift applied to the second playback stretches the original length of the sound. A digital constraint is implemented, which neutralises audio activation whenever the detected pitch is mapped to the same sound than the previous. As such, the density of the sonic construction depends on:
a) the intrinsic density of each audio sample,
b) the time length of the sample,
c) the loudness of the zither relatively to the detection threshold, and
d) the speed of my zither playing; usually I do not play loud at high speed, so as to leave space for musical details.

The previous section presented the adopted combinations of zither techniques and sample banks, and each creates a characteristic type of musical motion.

Combination # 1

With audio sample bank 1 the zither is dribbled and played with the bow[60]. As an input to the software it activates sounds of bass guitar, ocean waves, water drops, thunder and wind. The combination creates a changing continuum where the chromatic and textural variations from the zither merge into sounds of nature, submerging and emerging from the landscape. In the merging of acoustic and digital sounds, at times one shapes the attack, and the other shapes the release. Long-lasting bass guitar samples modulate the landscape with a dense sonic body. Their continuity allows for attention to focus on other streams of sound. But their appearance and disappearance attracts attention, creating points of higher intensity. Their disappearance is often associated with the resolution of a musical phrase, be it resolved with the zither, or the sound of a single water drop.

At times the music can become very dense and complex. Attention is then likely to shift away from details, as perception decreases resolution in order to embrace the whole. At lower perceptual resolution one perceives progressive continuities, where successive events seem to display a similar interval of motion, fulfilling the expectation that once something begins to move in a certain direction, it will continue to move in that direction (Gestalt of good continuation). When density and complexity then faint away, attention shifts to detail, and working memory increases perceptual resolution (perceptual resolution increases with attention, be it automatic or deliberate [e.g Knudson 2007]). At high resolution, ambivalent discontinuities become intense. The overall semantics of this combination are environmental, but it entails an expressive semantic dimension as well.

Combination # 2

With sample bank 2, the zither is played with hands, bottlenecks and pick.[61] The samples are from bass guitar, dobro (metal body guitar), and zither (played with metal pick and bottleneck). All samples have short durations and short attacks. Similarly, I play the zither with interruptions and silences. The combination of acoustic and digital sounds leads to an expressive pathos, as I avoid easy developments in the musical phrasing. In contrast with Combination #1, where the musical motion seems driven by nature, now the musical motion seems definitely human-scaled.

---

60 see small audio recording at http://adrianasa.planetaclix.pt/research/ArpeggioDetuning.htm#music
61 ibid.

Silence is intense if it emerges unexpectedly, and sounds gain intensity when preceded by silences. In both cases, the discontinuity attracts attention, leading to an increase of perceptual resolution. At times, a greater amount of complex musical phrasings makes for perception to decrease resolution, so as to embrace the whole. That is unlikely to last though, because one's attention is constantly being attracted by rapid musical developments. The overall semantics of the musical motion is expressive, as attention is drawn to the performer's expressivity.

Combination # 3

With sample bank 3 the zither is played with bow and bottleneck, activating piano notes and digital timbres.[62] The digital timbre samples are quite long, rich in bass, with gradual attacks and releases; their layering sustains sonic density. The body of the sound is inlaid with a wealth of surreptitious chromaticisms, unfolding at a textural level. The zither bow plays in unison, emerging from the sonic stream, so as to submerge once more. The piano samples - low keys - create soft punctuation.

The music unfolds at a slow pace, as if it were driven by forces greater than human time. There are no radical discontinuities. While sustained sonic density creates a space of overall continuity, conscious awareness is invited to focus upon ambivalent discontinuities, which makes chromaticisms and harmonics intense. At times attention may also shift away from details, and drift to internal states. The overall semantics of musical motion are more environmental than expressive.

The parametric model of the thesis needs to summarise the dynamic and semantic aspects of the work. It will visualise the overall range of sonic continuities/ discontinuities, and consider the semantic dimensions of the music together with the other semantic aspects of the work. It should also visualise how these aspects may change in combination with other musical idioms.

5.4 Extending Creative Strategies to the Image and Audio-Visual Relationship: AG#2

The creative strategies from the Arpeggio-Detuning were extended to a 3D audio-visual software, in which sound and image affect each other reciprocally. The Arpeggio-Detuning

[62] see small audio recording at http://adrianasa.planetaclix.pt/research/ArpeggioDetuning.htm#music

explored the principle of sonic complexity, expanding a musical language that emerged with the first version of my audio-visual instrument. The principle was considered at low level in the digital architecture. The new audio-visual software is called AG#2, and it considers the principles of visual continuity and audio-visual fungibility at a low level as well. Figure 26 outlines how the new, and final version of my audio-visual instrument works.

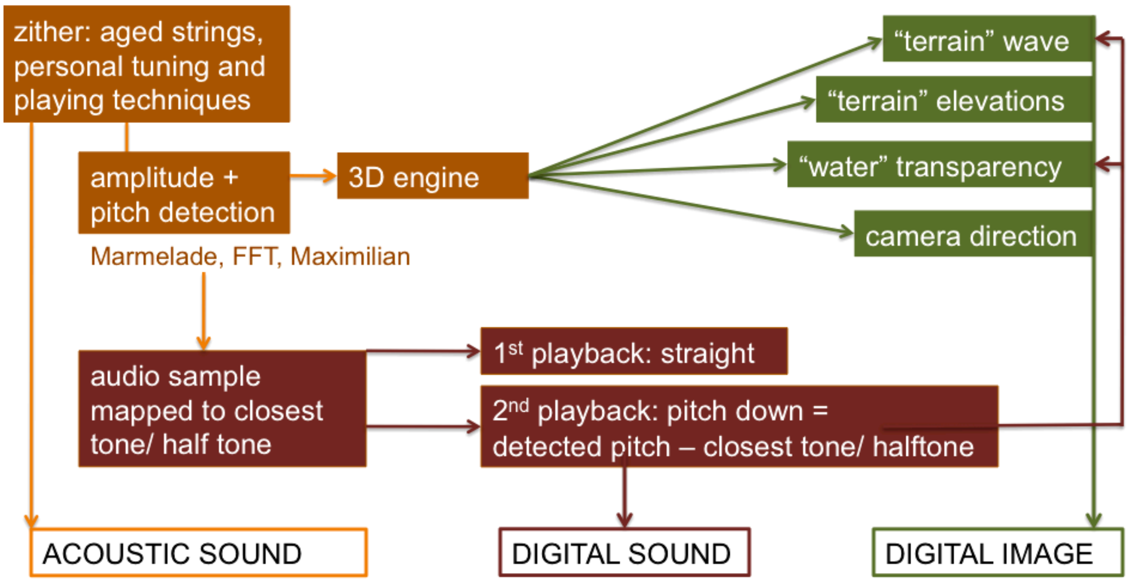


Figure 26: Technical diagram of the audio-visual instrument with AG#2

The 3D engine of AG#2 was written from scratch by John Klima, in C++. He implemented the engine in combination with the Arpeggio-Detuning, which uses the hardware abstraction layer *Marmalade*[63] and the audio library *Maximilian*[64]. I made the design specifications, the 3D world, the images applied to the 3D world, the audio-visual mappings and parameterisations. Furthermore, AG#2 has a configuration file, which can be understood as a threshold between instrument and composition. If we view the notion of instrument through that lens, the instrument alone does not guarantee compliance with my creative principles of sonic complexity, visual continuity and fungible audio-visual relationship. Thus I will describe the instrument and its configuration as a whole.

[63] Cross platform, iOs/Android development kit, https://www.madewithmarmalade.com/
[64] By Mick Gierson, https://github.com/micknoise/Maximilian

5.5 3D World and Visual Motion

Similarly to AG#1, AG#2 renders a 3D world, and the projected image is a shifting camera view over that world. But the visual dynamics are now much more organic, because the 3D engine is created from scratch, considering the principle of visual continuity.

The purpose of the image is to create an organic stage scene, without counteracting the musical motion. The principle of visual continuity dictates that to keep the music in the foreground one must dispense with disruptive visual changes, i.e. radical discontinuities. Those would automatically attract attention, and subordinate audition to vision in multisensory integration. There can be a wealth of discontinuities at a level of detail, as long as Gestaltist principles enable perceptual simplification. Visual dynamics can exhibit progressive continuities and ambivalent discontinuities. Recapitulating, with progressive continuities, successive events display a similar interval of motion, conveying Gestalts of good continuation. Ambivalent discontinuities are simultaneously continuous and discontinuous. At low perceptual resolution, one senses continuity. But if one deliberately pays attention, working memory optimises perceptual resolution causing discontinuities to become more intense.

In AG#2 the virtual camera shifts horizontally, creating smooth progressive continuities. The 3D world morphs upon audio input detection, in ways that take maximum advantage of the Gestaltist principle invariance: we recognize an object independently of rotation, translation, scale, elastic deformations, different lighting, and different component features.

The 3D engine enables the superimposition of several image layers, each with its own dynamics. By superimposing a set of images with many transparencies, I created a great deal of ambivalent discontinuities. Shadow and lighting effects also create ambivalent discontinuities. The image layers implement a GLES 2.0 real-time shader (software that runs on a graphics card, which determines how an object should be drawn), supporting four RGBA texture files (red, green, blue, alfa/ transparency). These files provide information about: a) lighting and surface effects, which establish the primary coloration and transparency of the 3D world element; b) the surface qualities primarily affecting the refraction of the light source, the shadows and lightings; and c) the reflection of the world around the element.

Figure 27 and Table 5 describe the 3D world and its dynamics.

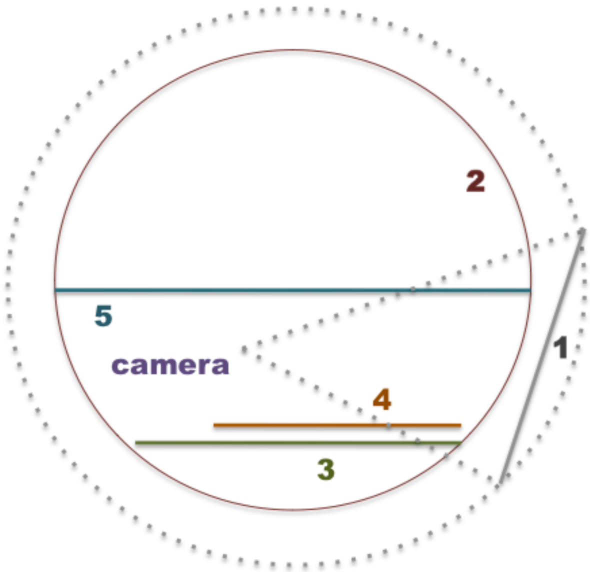
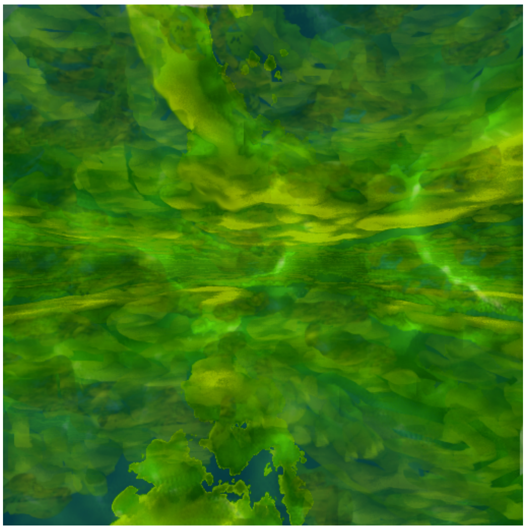


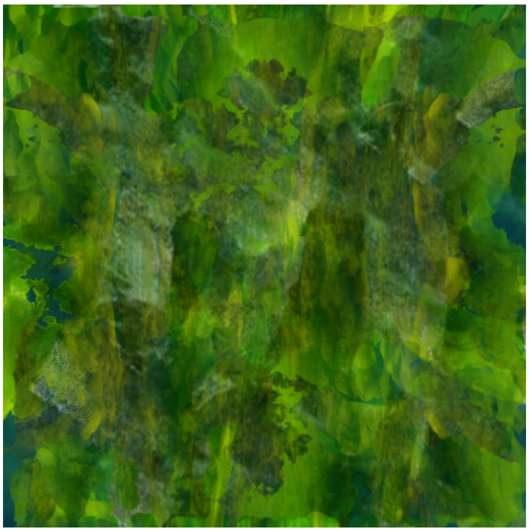
Figure 27: 3D world structure in AG#2 (cut)

1	**Background:** static image rendered full screen, the backdrop.
2	**Sky dome:** an enclosing, rotational image. It rotates consistently, independent from input sound, creating progressive continuity. Sky dome transparencies allow portions of the background image to become visible, creating ambivalent discontinuities.
3	**Environment terrain**: a large-scale 64x64 vertex mesh, which can be thought of as a distant "landscape". It is not affected by sound, and has no intrinsic motion. Yet the image applied to the mesh creates ambivalent discontinuities, as its transparencies allow portions of the sky dome and the background to become visible.
4	**Foreground terrain**: a 64x64 vertex mesh with an image applied, which can be thought of as the ground the 3D camera is standing upon. Audio input displaces individual vertices, creating elevations. It also produces undulations modulated by sine waves. These visual dynamics convey Gestalts of invariance. Image transparencies allow portions of the environment terrain, the sky dome and the background to become visible, multiplying ambivalent discontinuities.
5	**Water surface**: a small-scale 64x64 vertex mesh, which moves consistently, with progressive continuity. Given the 3D camera stands upon the foreground terrain, it provides “underwater” views; the water appears layered upon the sky dome and the background, letting them show through. Transparency changes convey the Gestalt invariance, and water transparency is mapped to sound.

Table 5: 3D world structure and visual dynamics in AG#2

Figures 28-37 show the images created for two slightly different 3D worlds; prior to performance I choose one or the other. Figures 38-40 show screenshots from AG#2. And photos 7-12 are from a performance.





Figures 28 and 29: Alternative background images (for 3D worlds #1 and #2), 512x512 pixels

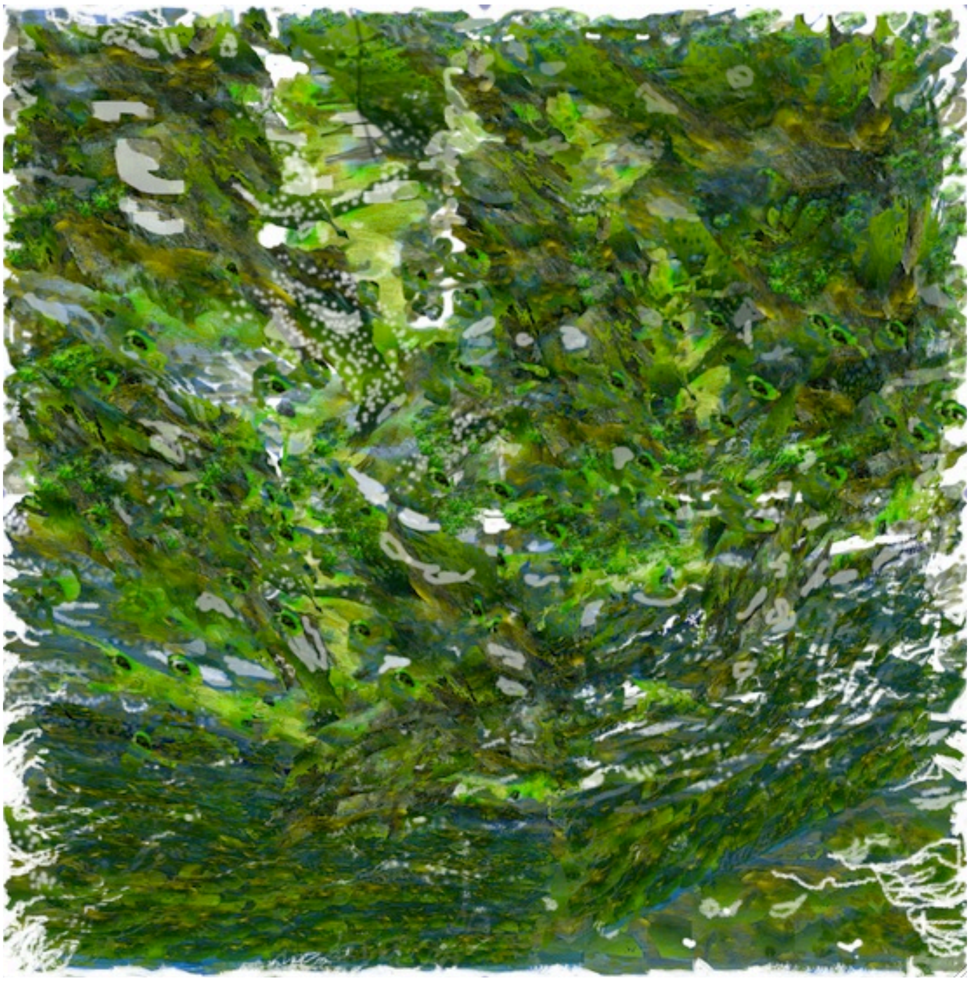
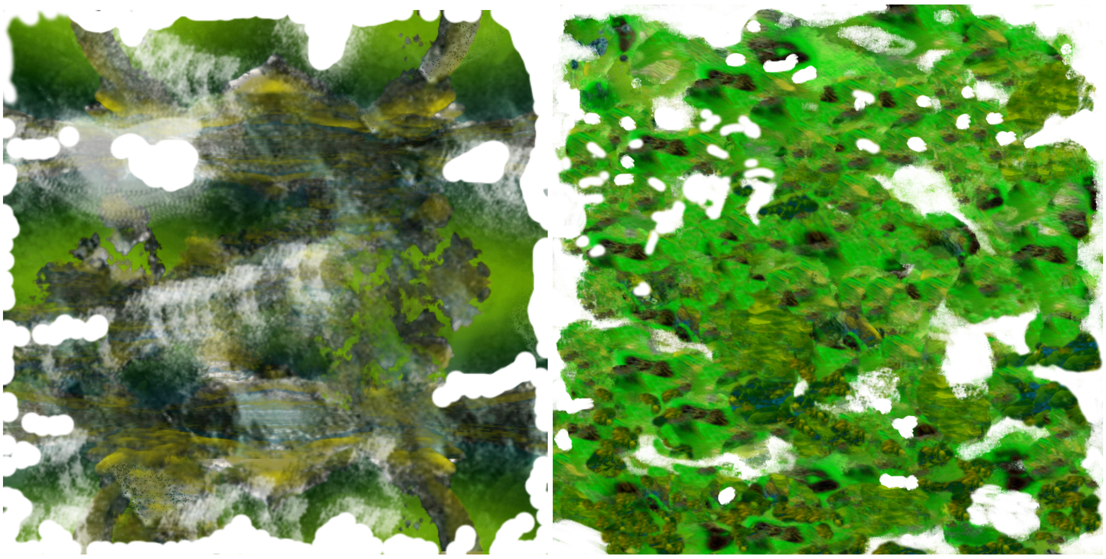
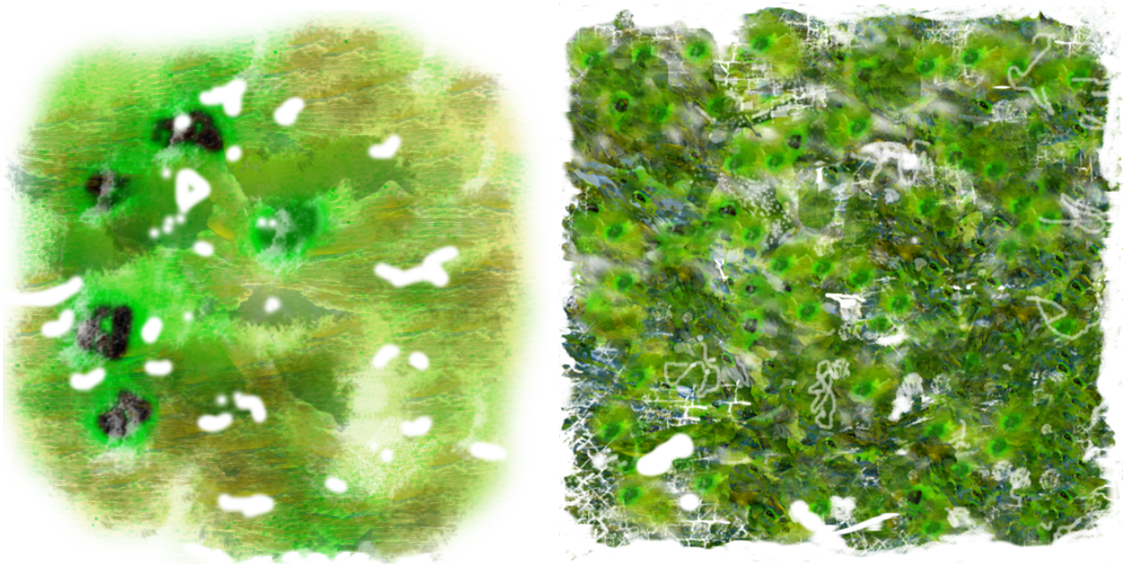


Figure 30: Sky dome image, 2048x2048 pixels



Figures 31 and 32: Alternative environment terrain images (for 3D worlds #1 and #2), 1024x1024 pixels



Figures 33 and 34: Alternative foreground terrain images (for 3D worlds #1 and #2), 2048x2048 pixels

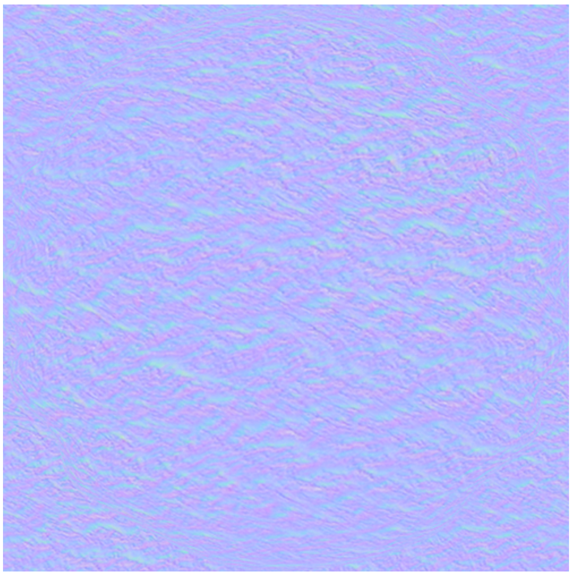


Figure 35: Image that establishes the surface qualities of the environment terrain and the foreground terrain (refraction of light source).
The RGB represents an XYZ vector indicating an angle and intensity of refraction. 1024x1024 pixels

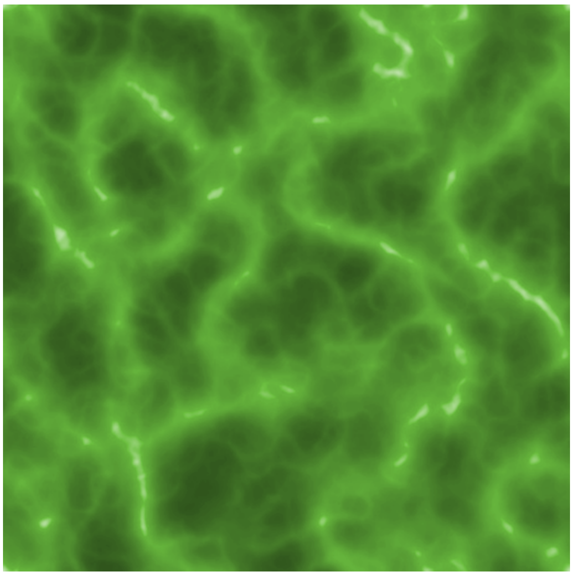


Figure 36: Image that establishes the primary coloration and transparency of the water; 512x512 pixels.

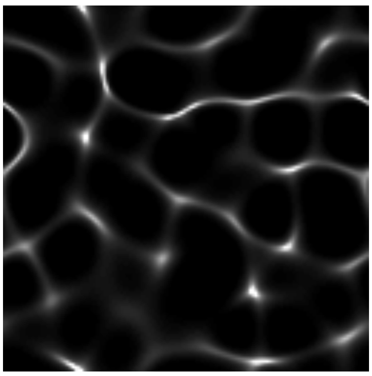


Figure 37: Image applied to the foreground terrain, describing discrete light events, such as refractions of light through water (stored in the environment map's alpha channel); 128x128 pixels.

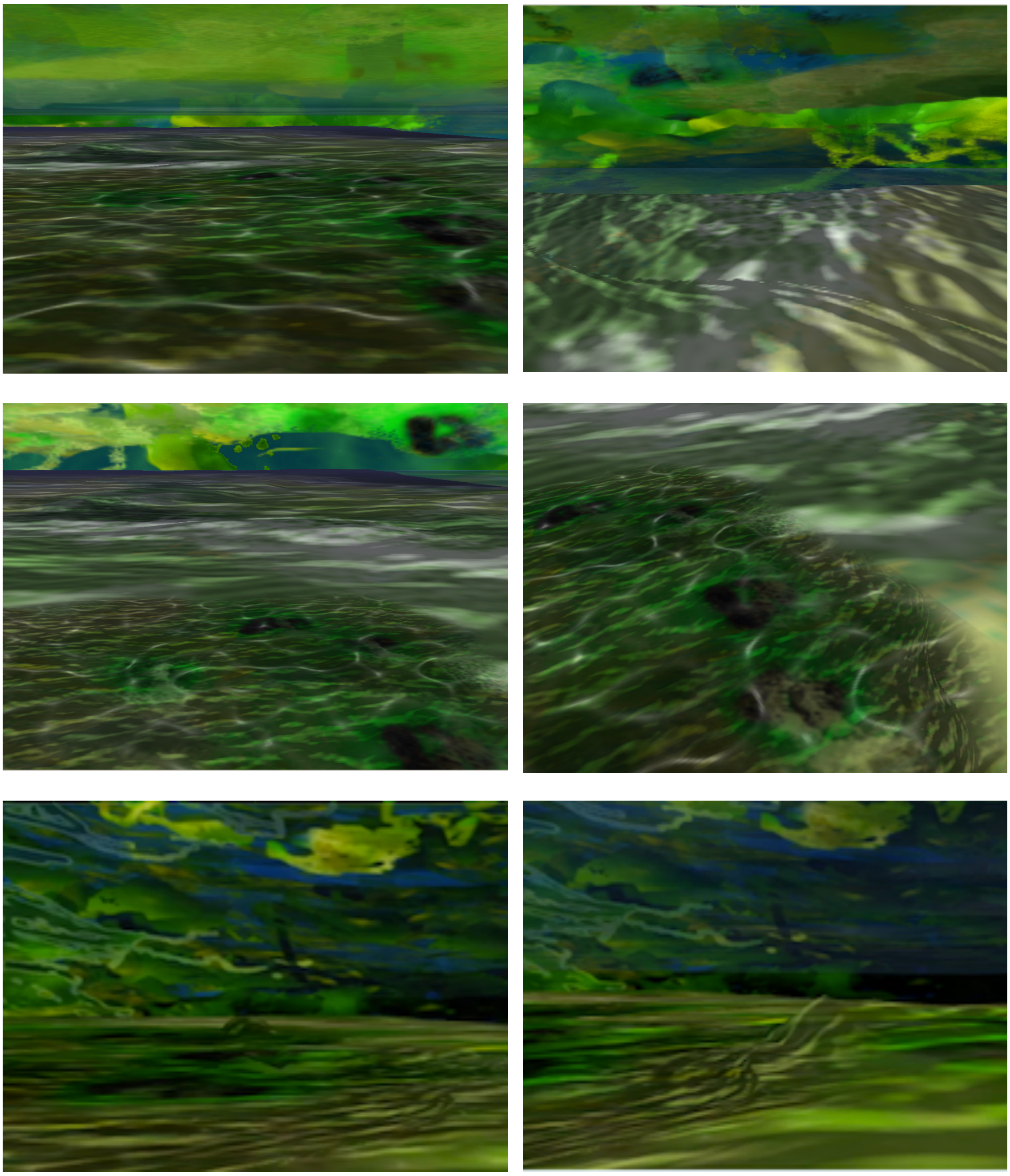
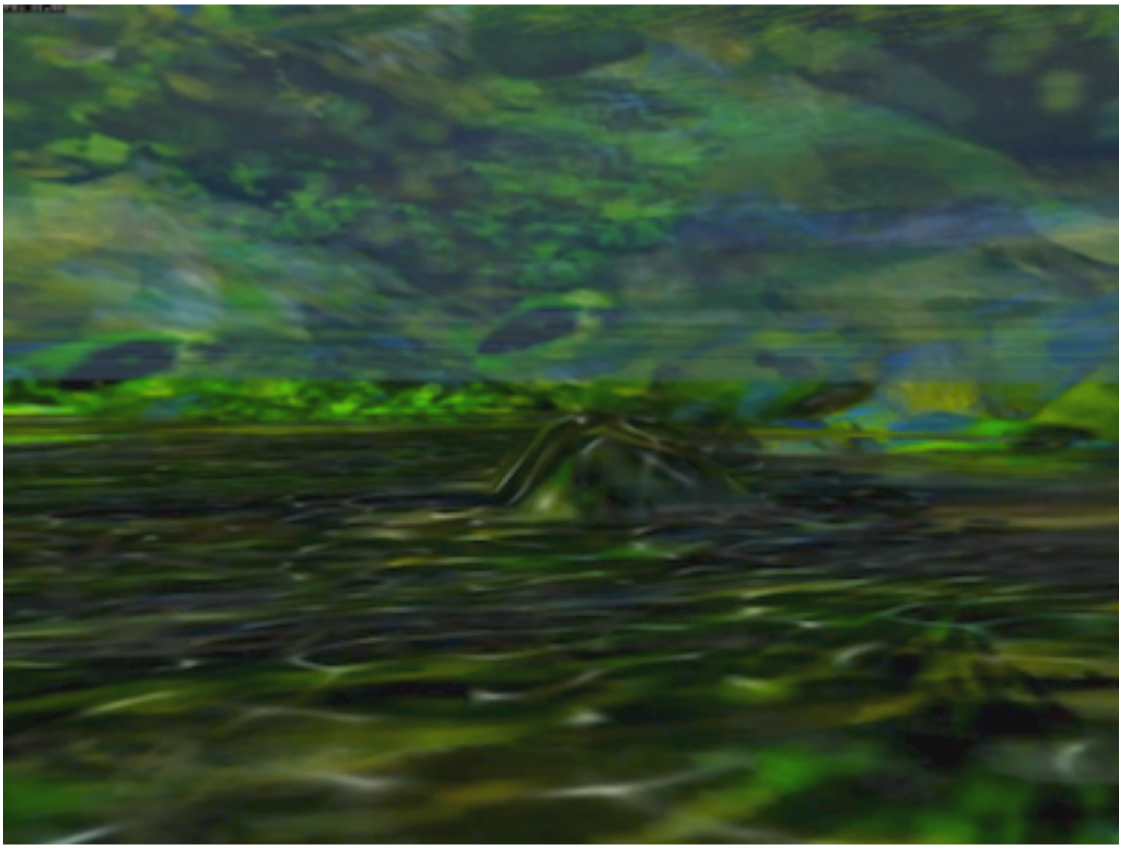
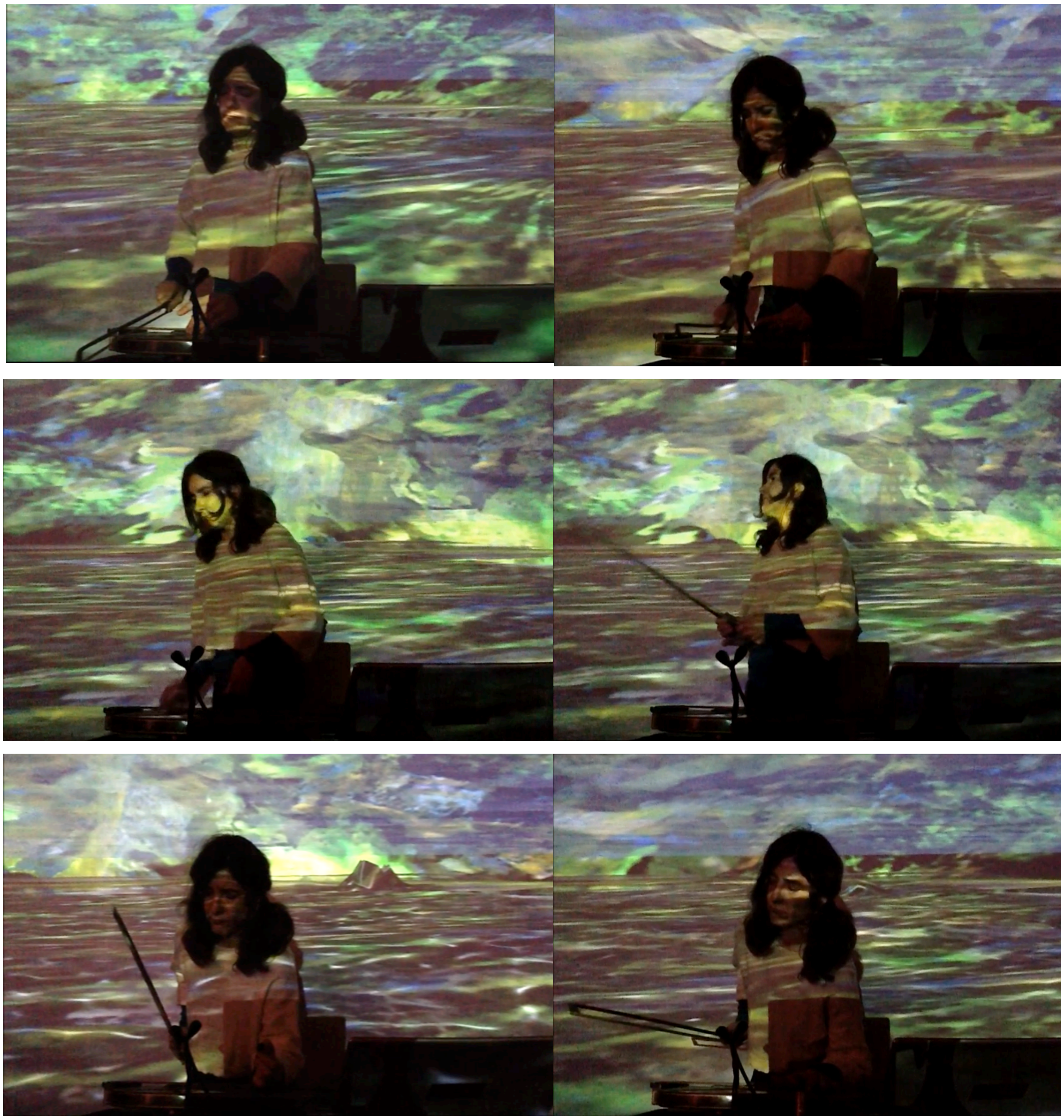


Figure 38: Screenshots from an early version of the 3D world made for AG#2



Figures 39 and 40: Screenshots from two recent 3D worlds, #1 and #2 (terrain elevations are being activated by audio detection; screen capture at http://adrianasa.planetaclix.pt/research/AG2.htm)



Photos 7-12: Performance with the final version of the instrument at SctrachBuilt Studios, Lisbon 2014 (video at http://adrianasa.planetaclix.pt/research/AG2.htm)

As the digital 3D world is projected over the performer the visual semantics are highly environmental, as happened with AG#1. The overlapping images with transparencies and the lighting & shadow effects invite attention to focus on a greater amount and variety of ambivalent discontinuities. Moreover, the visual dynamics place the threshold between ambivalent discontinuities and radical discontinuities in question.

As the audio input modulates the foreground terrain, occasionally an elevation becomes disproportionally high because the software does not calculate an actual number. Often this can be the result from a division by zero at some point in the very complex chain of calculations. However, after close scrutiny, and many error “traps,” it does not appear to be so. One could call this a “bug”: in theory, the analysis of the audio input should always produce a number. An actual number could be clamped to a given number interval, but one cannot do this with a non-number. Since AG#2 doesn’t “crash”, the “bug” is actually fascinating. At a purely conceptual level, it stresses how a digital “black box” might exceed the knowledge of any skilled low-level programmer. And at perceptual level, it raises questions about the interplay of automatic and deliberate attention in a large compositional structure.

A disproportional terrain elevation conveys the Gestalt of invariance: an object is recognised independent of its scale and elastic deformations. The question is if it also conveys the Gestalt of similarity, whereby similar elements are grouped as part of the same form. We still apprehend an elevation in the foreground terrain, which implies a level of similarity. Yet the elevation is much bigger, and the difference is clearly noticeable. So, to which extent does the work comply with the principle of visual continuity?

The answer is less straightforward than it was with AG#1. I never experienced radical visual discontinuities with AG#2, but cannot account for the audience - basically due to the different timescales of our experiences and our potentially different predispositions. During a performance I don’t look at the image; my assessment comes from the preparation work, as well as short performance videos. For example, the one at http://adrianasa.planetaclix.pt/research/AG2.htm#VIDEOS is 4’47” long, and shows a single disproportional elevation at 2’52”. At the timescale of a performance (about 35’) those events are much sparser. There are larger periods with repeated visual behaviours. That repetition can lead one’s visual resolution to decrease, in which case the nervous system should not respond in strong and automatic ways to any visual change. Alternatively, sustained attention to the image can make the visual details gradually more intense; one becomes more susceptible to automatic attention.

Another question is what happens after a disproportional visual change, when the terrain elevations return to their previous average heights. Whether the audience experienced a radical visual discontinuity or not, that bigger change is likely to make them pay more attention to the image and the audio-visual relationship. At the same time, the return to previous visual dynamics creates a sense of closure, hence they can equally focus on the sonic construction - whose dramaturgy is more complex, including many more points of high intensity. In any case, one might not really distinguish whether perceptual motion is driven by vision or audition, inclusively because salient sounds activate the visual cortex

[McDonald et al. 2013]. In other words, neither user-study results nor neuro-imaging results would allow for absolute conclusions about the intensity of visual changes.

There is simply no *absolute* threshold between ambivalent discontinuities and radical discontinuities, just like there is no *absolute* threshold between deliberate and automatic attention. The threshold depends on what we are looking at. It seems useful to recall Otto Roessler's description of knowledge as a "window" over reality [1996]; it was mentioned in chapter 3.2. Mandelbrot (an earlier Chaos scientist) looked for a method to compare multiple "windows", so as to describe a fundamental property of nature. He introduced the notion of level of description: the temporal interval within a time series, corresponding to the degree of resolution within the observation [1982]. He found a way of measuring and comparing different time series with respect to their temporal density, by assigning a fractal dimension to statistically scale-invariant spatio-temporal structures.

Our taxonomy of continuities and discontinuities could be seen as a type of "invariant structure", because it can be used to describe moment-by-moment situations as well as large compositional structures. However, assigning fractal dimensions to attention dynamics at different timescales is beyond the aims of this thesis. Our parametric visualisation model will rather represent the range of visual discontinuities according to my individual assessment, at the timescales of my experience.

The audience's assessment could differ from mine, but the disproportional terrain elevations don't seem problematic because the visual motion does not counteract the musical motion. This estimation is based on two factors. The first is that those occasional bigger visual changes never coincide with points of low sonic intensity, because the audio detection threshold is high. And the second is that the image will always convey Gestalts of invariance, unlike the musical dramaturgy, which unfolds through a set of different musical forms.

5.6 Audio-Visual Mappings

AG#2 has a fungible audio-visual mapping. Chapter 3.5 elaborated on how we are driven to form conclusive concepts of causation at the expense of ignoring diverging information. The study of chapter 3.6 demonstrated that with a fungible audio-visual mapping perception embraces convergences and divergences as inconclusive concepts. Recapitulating, a fungible mapping combines synchronised and seemingly non-related components, exhibiting complexity enough to be confusing. Conscious awareness does not segregate fitting information and non-fitting information, and as the brain seeks for a global ecological fit, the sense of causation extends to the mapping as a whole.

Both AG#1 and AG#2 have fungible mappings, but the one of AG#2 is far more sophisticated. Beyond synchronised and unrelated audio-visual components, now the mapping also includes components that sometimes synchronise, and other times not. The mathematical difference between the detected pitch and the closest tone/ halftone, which is applied to the processing of digital sounds, is also applied to the audio-visual mapping, as a delay value. Table 6 systematises how different visual parameters are mapped to different data streams from audio analysis.

Visual parameters modulated by sound	Data mapped to visual parameters
Synchronised with detection:	
Camera view (target direction)	Closest tone or halftone
Terrain elevation (target height)	Amplitude + difference between detected pitch and nearest tone/ halftone
Distance between elevation and 3D camera	Amplitude
Synchronised with the 2nd playback of each activated digital sound:	
Water transparency	Difference between detected pitch and nearest tone/ halftone
Delayed upon detection:	
Wave frequency applied to foreground terrain	Delay time = difference between detected pitch and nearest tone/ halftone; Frequency = nearest tone/halftone
Dynamic visual events independent from audio detection	
Sky dome rotation	
Shader applied to water surface (water reflexions)	
Shader applied to foreground terrain (surface effects)	

Table 6: The fungible audio-visual mapping in AG#2

The previous section described each of the 3D world components (Table 5): the static background image, the sky dome, the environment terrain, the foreground terrain, and the water surface. Every detected audio input places a virtual sound emitter in the world, upon the foreground terrain, at variable distances from the virtual camera. The sound emitters are visualised as terrain elevations, reaching variable heights. When input is detected, a digital sound is emitted, a terrain elevation emerges, and the camera view shifts toward a new target direction.

Although these three events are synchronised with detection, they do not appear consistent. First, because the digital sounds have very diverse attacks (time it takes to reach maximum amplitude); with short attacks they appear synchronised with detection, and with long attacks they can appear delayed. Second, because the detection threshold is high, which means that the acoustic sound may not affect the digital sound and image. And third, because at times the digital audio banks are switched off, which means that the acoustic sound solely affects the image. These inconsistencies create complexity, but they would not suffice to confound causation, as shown with the mapping prototype Interrupted Synchrony from the study of chapter 3.6.

The changes in water transparency are synchronised with the second playback of the digital sound mapped to the detected pitch. These additional points of synchrony reinforce one's sense of causation, yet the base cause-effect relationships are also more confounded. First, because these audio-visual events are not synchronised with the corresponding physical gesture and acoustic input sound; the delay time is equal to the duration of the activated digital sound. And second, because the changes in water transparency are sometimes not clearly noticeable.

The foreground terrain undulates according to variable frequencies, and frequency changes occur with variable delay times upon detection. The delay times are equal to the difference between the detected pitch and the nearest tone / half tone. Sometimes these values equal zero, because some zither strings are tuned to tones and half tones, creating points of synchrony. Hence, sometimes the undulations and the elevations in the foreground terrain happen at the same time, and other times not. Their mismatch makes the terrain behave in an organic, seemingly chaotic manner.

The study about audio-visual mappings (chapter 3.6) showed that variable delay times do not produce a sense of causation if points of synchrony are too sparse. That happens if the mapping is isolated, but not when it is combined with other, synchronised audio-visual components. As the fungible mapping fosters causal percepts and we remain uncertain about the base cause and effect relationships, we can also create perceptual links between sonic events and visual dynamics that do not depend on audio detection. That is the case of the sky dome and the water, which move consistently, at different pace and in different directions.

Recalling the perceptual effects of a fungible mapping, perception binds sonic and visual shapes that change or move simultaneously (Gestalt of common fate) or adequately proximal in time (Gestalt of proximity / sequential integration). Points of synchrony bring a sense of causation, but one remains unsure about whether the connection corresponds to a digital mapping, or if it is purely perceptual. As complexity confounds the base cause and effect relationships, perception cannot automatically discount sensory information according to previous concepts of causation. It embraces fitting and non-fitting information as inconclusive concepts.

Our parametric model can visualise the medium ecological fit of a fungible mapping in a very direct way, but the semantics created by the mapping merge with the global semantics of the work. The fungible mapping has an informational semantic dimension, derived from certain ecological fit: synchrony tells us that sonic and visual events are related through digital mappings. To represent this semantic aspect of the mapping, we need to consider how it combines with the semantics of sound and image, whose figurative qualities create informational semantics as well (see sections 5.3.4 and 5.5).

5.7 Sound Spatialisation

This section describes how AG#2 applies perceptual cues to digital sound spatialisation, and how it rules out undesired outcomes, considering the principles of sonic complexity and audio-visual fungibility.

In AG#2 the digital sounds are emitted through the elevations of the foreground terrain in the 3D world. The 3D engine implements 3D sound: the audio output depends upon the position of the elevations/ sound sources in the 3D world, relative to the 3D camera/ listener. If the camera/ listener or the elevation moves, the audio output reflects this shift. The thesis has previously explained how 3D sound differs from 2D sound (see chapter 2.2.3). 2D sound is characteristic of CDs, films and TV; each aspect of the sound environment is directed to a particular output speaker. Conversely, with 3D sound the sounds are dynamically routed to the speakers; sound spatialisation depends on visual dynamics. 3D sound can be used with stereo diffusion, as much as 2D sound can be used with multichannel diffusion.

3D positional audio effects apply the study of human perception mechanisms: one links digital sound and image because the software makes use of cues employed in the physical world [e.g. Dourish 2004]. Humans have a remarkable ability to localise sound sources accurately and rapidly. In the physical world, we use multiple cues for detecting the position of a sound source. 3D positional audio effects solely need to explore a few of those perceptual cues to provide a sense of verisimilitude. For example, both loudness and reverberation are cues for distance, but distance can be inferred from loudness alone. Recalling Snyder's notion of semiactivated memories, each event activates memories of similar events, so that every piece of conscious information embeds unconscious memories [2001].

The dominant paradigm for 3D sound is multichannel diffusion: the combination creates a high level of audio-visual fit, which makes the direction of causation clearly discernible. As Pressing observed back in 1997, this leads perception to prioritise vision over audition. Yet the audience's sense of spatial presence in the virtual world beyond the screen does not require a high level of audio-visual fit. For example, film often throttles the fit of sound and image, as observed in chapter 2.6.

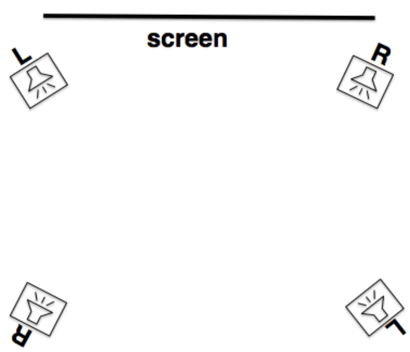


Figure 41: Inverted double-stereo system – a simple way of exploring the principle of audio-visual fungibility with digital 3D sound spatialisation

Chapter 5.2.5 described how the combination of 3D sound and inverted stereo diffusion conveys the principle of audio-visual fungibility (see Figure 41). The point is to combine audio-visual components that convey a sense of causation and components that do not, creating complexity so as to confound the base cause-effects. On the one hand, the speed at which the sounds move through the loudspeakers corresponds to the speed of visual dynamics, conveying causation. On the other, cause-effects are confounded because the audible sound does not really reflect the position of the sound emitter in the digital space. Inverted stereo diffusion would not interfere with people's conclusiveness about causation if synchrony were tight; perception would override the sound position, namely with the ventriloquist effect (perceptual dislocation of a sound towards a corresponding visual target [Pick et al. 1969]). Yet with a fungible audio-visual mapping, concepts of causation remain inconclusive.

3D sound spatialisation also respects to the principle of sonic complexity, because in performance I do have direct control over how the image affects the sound. This creates certain unpredictability, and it is necessary to rule out undesired outcomes. With the first version of my instrument, the process was a blind trial-and-error, because most spatialisation features were concealed in FMOD, the audio library used for AG#1. That is not the case with AG#2, where those parameters are specified and implemented from scratch.

5.7.1 Applying Perceptual Cues for Direction

Creating a volume difference between the outputs from each speaker suffices for an illusion of directionality and audible perspective. In the physical world, the ear further away from the sound source receives less sound energy due to the head's shadow [Toiviainen 2008]. A mono sound[65] can be spatialised by using two or more loudspeakers (e.g. a stereo system), so as to create the impression of various directions, as in natural hearing. The direction of the sound source is suggested by the relative amplitude of the signal sent to each loudspeaker.

AG#2 spatialises the digital sounds in this way, as video games usually do. If I used a single stereo system or a surround system, a visible sound emitter on the left of the screen would sound louder through the left speaker. With an inverted stereo system, the emitter on the left of screen will sound louder through both the front-left speaker and the back-right speaker.

The elevation of a sound source is detected less accurately than its direction on the horizontal plane [Toiviainen 2008]. The most important cues are related to the spectral shape of the perceived sound, as the sound spectrum is modified by reflections in the outer ear. With 3D software one can apply frequency filters so as to create spectral variations mapped to elevation. Flight simulation games usually do so, but not AG#2. One reason is that keeping the terrain elevations in the virtual camera frame requires both the camera and

[65] Using mono rather than stereo sounds is useful with 3D rendering software such as AG#2, since the processing of a mono sound requires much less CPU than the processing of a stereo sound.

the sound emitters to remain on the horizontal plane. The other, equally important reason is that a frequency filter would homogenise the frequency range of the digital sounds, and counteract the musical concept of pitching down the sounds according to the difference between the detected pitch and the closest tone/ half tone.

5.7.2 Applying Perceptual Cues for Distance

AG#2 explores two perceptual cues for the distance of a sound source. One is loudness: a sound placed at greater distance looses more energy before it reaches the ears than a sound placed nearby. The other is motion parallax: the translational movement of the listener causes larger azimuth change for nearby objects than for distant ones.

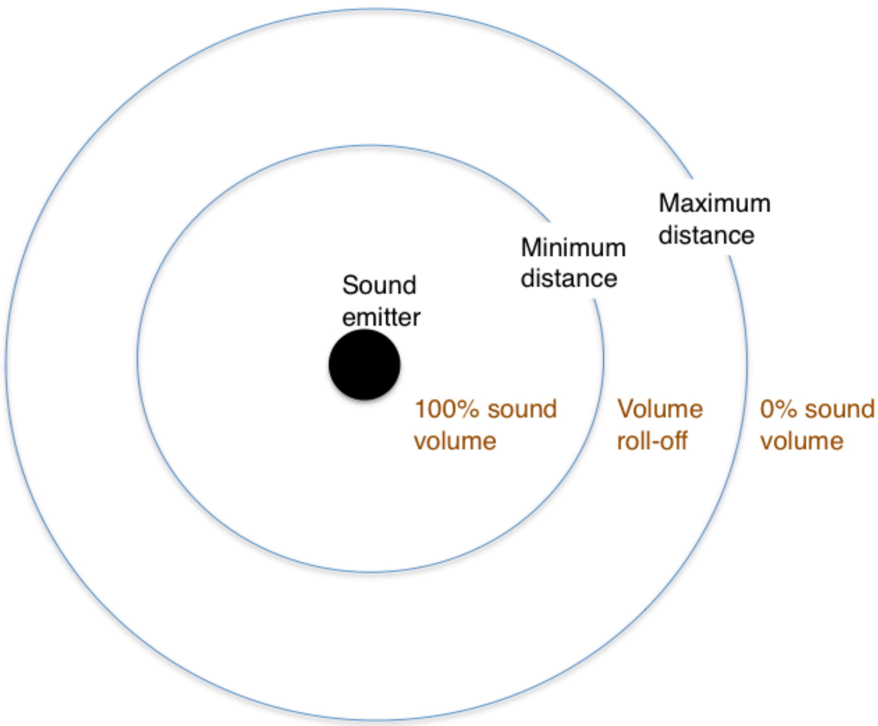


Figure 42: Area-based sound

Both these cues are commonly used in 3D software.[66] The sound emitters in a digital 3D world are area-based sounds, audible when the 3D camera/ listener is within a certain area (Figure 42). Usually there is peripheral volume attenuation, according to a *roll-off* algorithm. The sound begins to fade out at certain point called *minimum distance*, and it becomes

[66] The quality of a particular sound also plays a role in estimating distance; for example, shouting from a long distance can have a higher perceived loudness than whispering from a short distance [Toiviainen 2008]. Yet these kind of factors can hardly be quantified with mathematical functions.

inaudible at a point called *maximum distance*. The virtual camera position determines the volume of the sound and its azimuth changes; sonic dislocations are bigger when the camera is near the visible terrain elevation/ sound emitter, and smaller when it is distant.

When I'm playing music azimuth changes are welcome, but rapid, uncontrolled volume changes are undesirable. In AG#2, sound emitters closer to the camera/ listener are louder than those further away, but the sound areas are enlarged so that the activated digital sounds remain audible even at great distance. The space between the points of minimum distance and maximum distance is quite large, and volume decreases smoothly. Furthermore, the camera spins changing view direction, but it remains in the same spot (same Cartesian coordinates). If it dislocated through the 3D world, traversing sound areas would create rapid, uncontrollable volume changes.

5.8 Sound Diffusion Quality

This section describes technical experiences and choices regarding amplification. It also points out how these choices affect the performative arena and the semantics of the work, which reflect in the audience's sense of spatial presence.

5.8.1 Microphone, Compression and Reverb

My zither playing techniques have been developed using a specific piezo microphone (Photo 13) attached to the body of the zither. The piezo seems to cut high-end frequencies, which could possibly be resolved with a DI box: ultimately the piezo response depends on voltage, and on what it is plugged into. In any case, I have tried several other microphones throughout the years, including Shadow contact microphones and expensive piano microphones. Yet this cheap, hand-made piezo[67] brings a particular timbre. It is quite responsive at low frequencies, and as it is not isolated in a box, it also captures the room, which gives greater body to the sound. These qualities tend to create feedback, conditioning volume levels; but within certain limits, feedback creates a wealth of overtones, compensating the piezo's high-frequency cut.

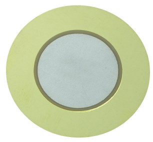


Photo 13: Piezo microphone

[67] http://www.conrad.com/ce/en/product/710521/KEPO-FT-41T-10A1-478-piezo-element-800-1-05-kHz-41-mm?queryFromSuggest=true

The piezo connects to a dynamic compression pedal (MXR Super Comp, Photo 14). Compressors clamp the dynamic range of sounds, by increasing the amplitude (volume) of low-amplitude sounds, and decreasing the amplitude of high-amplitude sounds. The compressor gives presence to subtle acoustic variations and overtones. As it decreases amplitude peaks, it also diminishes problems with feedback.

Photo 14: Compressor

The piezo signal outputs to the compressor for amplification, and also directly to the iPad, for signal analysis. Audio detection would happen too frequently and consistently if the input sound were compressed.

Returning to the audible zither sound, overtones and harmonics gain a greater and warmer presence when some reverberation is added. The compressor outputs to a reverb pedal, the Electro Harmonix Cathedral[68] (Photo 15), which sounds like an analogue device. Reverberation, or reverb, is a re-echoing of sound, perceived as resonance. Natural reverb exists everywhere, as sound is reflected by matter. Artificial reverb adds reverberation independent from the actual physical room. I use a subtle amount, so that it sounds like a spatial quality, rather than an effect.

Photo 15: Reverb

5.8.2 Testing Speakers and Amplifiers

The digital sound requires two stereo systems, but one is enough for testing sound quality. In my personal studio I have a full range stereo system (Behringer Truth B2031A), which I used for editing the digital sounds in combination with my zither techniques. I find this system quite satisfactory because it provides a very linear frequency response (from 50 Hz to 21 kHz). John Klima has a recording studio, where I tested another stereo system and several amplifiers.

Firstly I used a stereo system, JBL 4312A, to amplify both the acoustic and the digital sound. Those are full range speakers, yet they seemed less responsive at low frequencies than my Behringers. That is particularly problematic because many digital sounds are rich

---

68 http://www.ehx.com/products/cathedral

in bass, and their pitch down can produce extremely low frequencies. I compensated that with a subwoofer, an additional speaker dedicated to low frequencies. The digital output sounded very well, but still, there was an overall lack of definition in the zither sound, cancelling the complex wealth of chromatic variations. Hence, I decided to amplify the zither separately from the digital sound.

I tested the zither with different guitar amplifiers. The Carvin Valvemaster (4x12 cabinet), a big transistor amplifier for *heavymetal* guitar, was rapidly ruled out, as the zither sound lost its acoustic quality, and overall definition. There was another big transistor amplifier for guitar: a Fender Frontman 212R, the modern version of the venerable Fender Twin Reverb. It gave a good body to the zither sound, and a good presence to overtones. But still, the zither sounded more electronic than acoustic. There were also two smaller amplifiers: a Fender Frontman 25R, for electric guitar, and a Trace Acoustic TA50R, for acoustic guitar. The problem with those is that they “clean” the sound, suppressing certain qualities and overtones that are crucial to the timbre of the zither. I also tested the zither with bass amplifiers: a smaller one, Fender Rumble 30, and a bigger one, Ampeg SVT 3 Pro with Genz Benz (2x10 cabinet). Medium and high frequencies did not sound very bright with the Fender, but the Genz Benz gave a very good body to the zither sound; moreover, it permits equalization for overtone enhancement.

At some point I realized that the zither needed a valve amplifier rather than a transistor amplifier. I bought a Mesa TA15 (Photo 16), which is fairly portable. It can connect to any cabinet; I acquired a Warick 12”, but the bigger the cabinet, the fuller the sound. In any case, the Mesa amplifier shapes the zither sound with incomparable definition, body and warmth.

Photo 16: Amplifier and cabinet

It became clear that the zither sounds best through this valve amplifier, placed next to me on stage. In addition, it needs re-amplification through the stereo systems placed around the audience, for two reasons. Firstly, because the amplifier cannot be very loud due to feedback. And secondly, because the acoustic sounds should merge, submerge and emerge from the digital sounds.

The overall improvement in sound definition enhances both the expressive and the environmental semantics of the work. The Mesa amplifier on stage reinforces the performer’s presence, improving the expressive dimension of the work (corresponding to

Emmerson's 'local' functions [2007] and Ciciliani's 'centripedal' performance tendencies [2014, 2015]). Given the amplifier, the volume of the zither through the speakers placed around the audience can be lower, and merge with the digital sounds without loosing presence. That makes the digital sound gain definition as well, improving the environmental dimension of the work (corresponding to Emmerson's 'field' functions [2007] and Ciciliani's 'centrifugal' tendencies [2014, 2015]). The amplifier on stage could potentially reinforce the informational dimension, but the acoustic and digital sounds can still be confounded, because the double stereo system outputs them both.

5.9 Chapter Summary and Conclusions

The chapter described an audio-visual instrument and performance language, which explore the creative principles of sonic complexity, visual continuity and audio-visual fungibility. Informed by the perceptual approach and the parametric model of the thesis, my practical work evolved immensely. Subsequently to describing the first version of the instrument, the chapter described how my creative strategies expanded in depth. That expansion started with a focus on sound organisation, followed by visual dynamics and audio-visual mapping, digital sound spatialisation and sound diffusion.

After creating the first version of my audio-visual software, AG#1, my musical strategies evolved with audio software, the Arpeggio-Detuning. Operating based on amplitude and pitch detection from the zither input, the software explores divergences between two data streams from pitch analysis: the detected frequency, and the closest tone or half tone. The difference between these two values is used as a pitch-down value, which applies to pre-recorded sounds.

As the chapter introduced the adopted combinations of zither techniques and digital sounds, it also proposed the term figurative sounds to describe recognisable sounds. Figurative sounds have informational load, which can convey expressive or environmental semantics. I described musical forms and digital constraints using the taxonomy of the thesis, which relates attention, intensity and continuities/ discontinuities; my analysis revealed that the semantics of music are inextricably related to the interplay of continuities and discontinuities.

The creative strategies from the Arpeggio-Detuning were extended to the final audio-visual software, AG#2. AG#2 uses a custom 3D engine, which allows for the principles of sonic complexity, visual continuity and fungible audio-visual relationships to be considered at low level in the digital architecture. The difference between the detected pitch and the closest tone/ half tone, which applies to the processing of digital sounds, is also applied to the audio-visual mapping.

The 3D world of AG#2 conveys environmental semantics, similarly to the world of AG#1. Yet the visual dynamics of AG#2 are more organic; the principle of visual continuity inspired me to maximise visual changes at a level of detail. The chapter described the 3D world, including its dynamic architecture and images, whose overlapping transparencies create progressive continuities and ambivalent visual discontinuities – with no radical visual discontinuities, nor polyphonic visual movements.

The principle of audio-visual fungibility led me to maximise the audio-visual mapping's complexity: in AG#2 the mapping combines not solely synchronised and unrelated components, but also partially related components. The difference between the detected pitch and the closest tone/ half tone is applied to the mapping as a delay value, and zero values also create points of synchrony.

Regarding sound spatialisation, specifying software from scratch enabled me to explore digital 3D sound in ways that consider the principle of sonic complexity. I described how AG#2 explores perceptual cues for direction and distance with a double, inverted stereo system, considering the principle of audio-visual fungibility. The research also focused on physical components, gleaning how sound quality could be improved so as to stress the distinguishing characteristics of the zither and the digital sounds. I also assessed how those improvements should affect the overall semantics of the work.

The chapter described my practical work in an integrated manner, considering that the audients' sense of spatial presence is influenced by the qualities of sound and image, the audio-visual mapping and the physical setup; which in combination shape a dynamic performative arena.

CHAPTER 6

Discussion

6.1 Chapter Introduction

This final chapter discusses the contributions of this research, including the relation between my practical work and the proposed parametric visualisation model.

Section 6.2 discusses how the model parameterises sonic expression, sensory dominance and spatial presence, considering the creative principles explored in this research.

Section 6.3 uses the model to represent my creative thinking with respect to instrument design and solo audio-visual performances. By extension, it points out how using the model to parameterise one's creative planning can lead to new questions about the audience's experience in large compositional structures.

Section 6.4 parameterises the work's versatility in collaboration with other musicians, based on four case studies.

Section 6.5 provides a summary of contributions.

6.2 Discussing Parameters

This section discusses how the model of the thesis parameterises sonic expression, sensory dominance and spatial presence, considering my practical work and the creative principles of sonic complexity, visual continuity and audio-visual fungibility.

6.2.1 About the Parameterisation of Sonic Expression

Talking about musical expression requires one to look at how interaction informs the dynamics and semantics of music. Our parametric model visualises those three aspects, and one can use it to discuss how one aspect influences the other, in any particular work.

As explained in chapter 4.3.2, from the dynamics and semantics of the music we cannot infer whether the work under discussion converges or diverges from my creative principle of sonic complexity. Indeed, the principle admits any type of continuity/ discontinuity, and different semantic dimensions. Conversely, real-time effort in sound organisation is a distinguishing characteristic of the principle. The reader might question if it is an appropriate parameter to characterise interaction, particularly considering other parametric visualisation models, which focus on interaction (Birnbaum's and Magnusson's). It is a choice, but a thoughtful choice.

The model of this thesis does not aim to replace those previous models, as it parameterises interaction at a higher level of abstraction. The model is very useful to frame the expansion of my creative strategies, but it would not be useful to visualise what that expansion entails. With the final version of my instrument, the interaction is far more engaging than with the first. Magnusson's model, which emphasises the epistemic dimension of human-computer interaction, is adequate to visualise some differences between the first and the last version of my instrument. His model does not aim at representing interaction with acoustic instruments, but if we consider the acoustic and digital components of the instrument as a whole, his model can represent how the interaction design evolved (Figure 43).

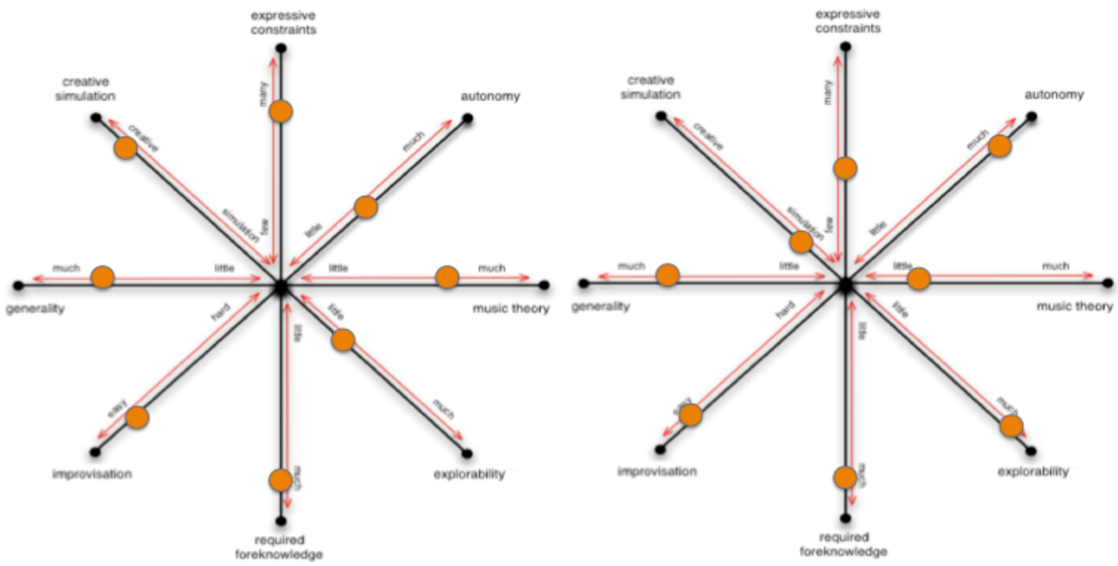


Figure 43: Using Magnusson's model to represent interaction with the first version of the instrument (left) and the final version of the instrument (right)

The *Foreknowledge* required to play the instrument is similar – playing the zither requires foreknowledge (the software does not require much because it already contains it). Given my immediate control over the acoustic output, the adaptability to novel musical situations is similar as well; hence the parameters *Generality* and *Improvisation* are fairly high.

AG#2, which implements the Arpeggio-Detuning, has more *Autonomy* than AG#1; I have less real-time control over the digital sound output, which is far more complex. But the output is equally conditioned by a greater amount of encapsulated musical knowledge. That knowledge involves music theory, but a theory that is not characteristic of any specific culture. There are more musical possibilities, hence less *Expressive Constraints*. Also, *Creative-Simulation* is lower in AG#2 than in AG#1, because the audio processing strategies are far more sophisticated.

My creative principle of sonic complexity can be discussed with respect to Magnusson's parameters. However, if we use his model to discuss my instrument we see that most graphically opposed parameters are actually not in opposition. *Explorability* has increased and *Creative-Simulation* has decreased, but that is the only opposition here. *Autonomy* increased, but *Improvisation* did not decrease. *Generality* remained the same, and *Music Theory* decreased. And *Expressive Constraints* decreased, whilst *Required Foreknowledge* decreased. These parameters are possibly in opposition with digital instruments, but there might be no opposition when the instrument combines acoustic and digital components. In fact, Magnusson's model is too detailed to reveal the notion of sonic expression formulated in this thesis.

Chapter 3.3 elaborated on this notion. It described how the interaction with a musical instrument entails not solely embodiment, but also a separation between performer and instrument, which reflects in expressive effort. This two-folded, reciprocal interaction informs the principle of sonic complexity, which dictates thresholds between control and unpredictability, calibrated so as to convey the performer's idiosyncratic expression. Parameterising embodiment/ separation does not seem very useful to the model of the thesis, because it does not provide any indication about what causes that separation. Parameterising control and unpredictability does not seem appropriate either, because the term unpredictability can be confusing. For example, volatile, digital interface behaviours may not be anticipated in real-time, yet they are determined by the code. And if we think of live coding performances, code may produce anticipated sonic results, yet there is certain unpredictability in terms of timing, as the performer needs to type the code (e.g. see Magnusson's contribution in [Sa et al. 2015]).

In our model, the parameter *Real-Time Effort in Sound Organisation* represents to which extent the sonic results depend on the performer's real-time effort, motor and/ or conceptual; effort is considered relatively to cognitive demand and musical timing. Little effort means that the music does not depend on the performer's real-time interaction, or the interaction is linear and clearly perceivable. In any case, little effort indicates a divergence from the principle of sonic complexity. High effort means that the interaction with the system does not feel immediate, and/or the system does not rule out undesired outcomes. My creative principle admits medium-high effort, but not maximum effort; maximum effort means that the performer does not control the construction of musical time. The interaction with my instrument entails medium effort, implying complex interface behaviours. The performer needs particular skills to play the instrument, yet a sense of immediacy conveys musical timing, and digital configurations rule out undesired outcomes.

As a parameter, *Real-Time Effort in Sound Organisation* is adequate for representing the creative principle of sonic complexity, and representing the interaction with any musical system or instrument. It seems a clear way of evoking the present notion of sonic expression, which includes, but is not restricted to, ratios of control and unpredictability.

6.2.2 About the Parameterisation of Sensory Dominance

The artistic debate about sensory dominance lacked resolution, and clarifying the problem is a key contribution of the thesis. That debate is not in contradiction with neuroscience and psychology: in many aspects vision tends to dominate over audition, and perceived causation is a relevant factor. However, attention can be manipulated so that one sense dominates over the other [Sinnett et al. 2007]. The thesis gleaned how an audio-visual performance work can foster the experience of sonic complexity, and simultaneously create a surplus of meaning through the audio-visual relationship.

So far, the artistic debate about visual dominance focused on the audio-visual relationship [Schaeffer 1966, Pressing 1997, Stevens 2009]. The thesis showed that visual dynamics are equally important. Sudden discontinuities attract automatic attention [Knudsen 2007], as our instinct is to detect causation in good time [Calvert et al. 2004]. My taxonomy of continuities/ discontinuities distinguishes between apprehensions automatically driven by stimuli, and apprehensions under individual control. This is very useful to assess whether the visual dynamics attract automatic attention, subordinating audition. Radical visual discontinuities do so, but ambivalent discontinuities, progressive continuities and steady continuities allow for attention to focus on the relations between the sounds.

The threshold between ambivalent and radical discontinuities must be evaluated with respect to specific audio-visual behaviours and related creative concerns. A user-study would need to take into account that salient sounds can affect vision without one being consciously aware of the interaction between the senses. It would also need to consider that any attempt to distinguish if attention is driven by stimuli or under individual control would condition the participants' perceptual experience. In any case, intensity depends on stimuli, panorama and perceptual resolution, which means that the threshold can't be considered *per se*. Whether it is defined based on the artist's experience or the audience's experience, it will always reflect a particular timescale, because over time, the same behaviours can appear more continuous or discontinuous.

Clearly perceivable cause and effect relationships are as problematic for the music as radical visual discontinuities. Highly congruent audio-visual relationships convey perceptual chunking. They are not demanding in terms of cognitive workload, but in the integrated audio-visual representation, perception prioritises fitting information over non-fitting, and vision subordinates audition [Pressing 1997, Boltz 2004, Kubovy and Schutz 2010]. Vision does not require maximum audio-visual fit to dominate over audition. It does not attenuate auditory input, yet visual dominance occurs as soon as we conceptualise causation [Spence 2009, Robinson et al. 2015]. I proposed a related notion of high ecological fit, so as to characterise every audio-visual relationship that conveys conclusions about causation. This notion of high fit embraces inconsistent audio-visual relationships, because inconsistency does not prevent conclusive concepts, as shown in Gestalt psychology and in the study from chapter 3.6.

Highly incongruent sound/ image pairings lead to separate representations for each sensory modality [e.g. Boltz 2004, Kubovy and Schutz 2010]. Yet they produce no convincing sense of causation, which is inconvenient if one aims the audience to sense some mechanical relation between sound and image. Also, binding incongruent sounds and images requires great cognitive processing effort; the audience can possibly not fully experience complex sonic constructions, given the limited capacity of short-term memory.

I defined medium ecological fit by establishing that an audio-visual relationship exhibiting medium fit produces inconclusive concepts of causation. Such inconclusiveness loosens the perceptual hierarchy, so that conscious awareness does not segregate fitting and non-fitting information. Moments of sensory unison convey chunking, but as attention is driven to discontinuity, we form separated auditory and visual representations. Accordingly, I introduced the notion of a fungible audio-visual relationship, which combines synchronised and non-synchronised components, exhibiting complexity so as to confound the base cause-effect relationships. The study from chapter 3.6 demonstrated how fungibility leads perception to embrace fitting and non-fitting information with inconclusive concepts.

The parametric model of the thesis enables us to infer sensory dominance from two parameters: *Visual Discontinuities* and *Ecological Fit*.

The axis *Visual Discontinuities* represents the range of visual continuities/ discontinuities at a given timescale. It is useful for several purposes. Firstly of course is the purpose of revealing whether the dynamics of the image convey visual dominance, in case that is a creative issue. The second purpose is inextricably related: to visualise whether the work under discussion converges with my creative principle of visual continuity. The third purpose is equally related to the previous two: to visualise the relation between the visual dynamics and the sonic dynamics, so as to provide a cue about whether sensory dominance is relevant or not. For example, the analysis of Steina Vasulka's *Violin Power* performance (chapter 4.4.1) shows that radical visual discontinuities can have important dramaturgical functions. Also, the analysis of Ryoji Ikeda's *Superimposition* (chapter 4.4.2) led to the conclusion that sensory dominance is not relevant in that work because the logics of sound and image are one the same; in other words, it would make no sense to highlight the relation between the sounds themselves.

The parameter *Ecological Fit* represents the audio-visual relationship. It includes two axes, and the global fit can be inferred from their combination. One axis is for the system output; it represents the audio-visual mapping and the speaker setup. The other is for the relation between physical gesture and system output. The global fit corresponds to the middle point of an imaginary segment uniting two dots, one on each axis. It does not require direct graphic representation because the middle point between two dots is straightforward; I opted for two axes so as to provide more cues about the work.

Low fit means that the sound/image pairing produces no convincing sense of causation. Binding is weak, as happens with interrupted random delay including sparse points of synchrony (see study from chapter 3.6). High fit means that the audio-visual relationship produces conclusive concepts of causation, and perception prioritises fitting information over non-fitting. This is the case not solely with consistent synchrony, but also with interrupted synchrony (see study). It is also the case when the changes in one sensory domain are always clearly preceded by changes in the other domain (see the analysis of

Vasulka's performance, chapter 4.4.1). Finally, medium fit means that the audio-visual relationship conveys concepts of causation, but one does not understand the base cause and effect relationships: perception binds auditory and visual stimuli, but not when they are under attention.

Medium fit indicates the work's convergence with my creative principle of audio-visual fungibility: it denotes sensing causation without understanding the base cause-effect relationships. Having two sub-parameters for ecological fit is useful because it enables a distinction between the system output and its relation with physical gesture. My practical work explores a fungible mapping as well as a fungible audio-visual relationship in space. Chapter 4.4 showed that cause-effect relationships can also be confounded in spite of constantly synchronised audio-visual mappings; namely if the performers' physical gestures seem unrelated (as in Ikeda's *Superimposition*), or if understanding cause-effect relationships requires understanding complex symbolic systems (as in Magnusson's *Threnoscope*). Furthermore, the parametric model can provide cues about why an artist may rather not create a fungible audio-visual relationship. For example, when physical gestures are not visible (as in Vasulka's *Violin Power* for Nam June Paik), clearly perceivable cause-effect relationships indicate an audio-visual instrument; they prevent us from thinking that the work is a film.

An audio-visual performance work conveys visual dominance if it exhibits radical visual discontinuities, high ecological fit, or both. Audition can be in command in any other case. The parameterisation of sensory dominance is useful whether that is the goal or not.

6.2.3 About the Parameterisation of Semantics and Spatial Presence

When analysing a work with respect to semantics and spatial presence, we must consider the physical and the psychological space of the work. In our parametric model, the *Performer Visibility* axis represents the performer's physical position. The axis includes three discrete points: *Integrated* means that the image functions as a stage scene, *Separated* means that the image is separated from the performer, and *Hidden* means that the performer is not visible. The axis *Performative Arena* corresponds to a higher-level parameter, which depends on the qualities of sound and image, the audio-visual relationship and the physical setup. It summarises information represented in other axes, and provides additional cues about elements that have no direct representation, such as the placement of speakers and the lighting.

The model uses three discrete points to represent three types of performative arena: *Local*, *Distributed* and *Extended*. These are not mutually exclusive; each denotes a particular type of semantics, and a work can overlap different semantic dimensions. The model extends Pressing's semantic characterisation of sounds [1997] to the audio-visual domain: it parameterises *Expressive, Environmental* and *Informational Semantics*.

The *Local* performative arena denotes *Expressive* semantics, i.e. a focus upon the performer's personal states and skills. It relates to Emmerson's 'local functions' [2007] and Ciciliani's 'centripetal' performance tendencies [2014]; both denote a focus upon the performer, and sound sources placed next to him or her, on stage. The notion of *Local Arena* implies that the qualities of sound and image influence spatial presence as much as the physical setup. For example, the audience can be led to focus upon the performer when the sonic construction reveals thresholds between her control over the system, and the system's unpredictability. Also, a projection over the performer can work as a spotlight (for example if he or she stands in the middle of a big yellow dot). Furthermore, the *Local* arena implies *Expressive* semantics but the inverse is not necessarily true. For example, there is no local arena in Vasulka's *Violin Power* because the performer is not visible; yet the violin timbre and the audio-visual fit draw a focus upon her skills, creating expressive semantics (see chapter 4.4.1).

The *Distributed* performative arena denotes *Environmental* semantics, i.e. a focus upon the environment. It relates to Emmerson's 'field functions' [2007] and Ciciliani's 'centrifugal' performance tendencies [2014]; both denote a focus on the environment where local activity may be found, implying sound sources distributed in space. Again, our model considers that the intrinsic qualities of sound and image influence the audience's sense of presence as much as the physical setup. For example, drone music deviates attention from the performer by creating an immersive environment. And a colourful image projected over a performer dressed in white can diminish their protagonism in relation to the stage scene.

The *Extended* performative arena represents the audience's mental transportation beyond the physical performance space. Such transportation requires *Informational* and *Environmental* semantics, whether these semantics come from the sounds, the image, or the audio-visual fit. When the sound or the image evokes a particular environment, one feels present in that space. And perceptual cues applied to an audio-visual relationship can covey spatial presence in a virtual world beyond the screen. The *Ecological Fit* parameter should indicate if that is the case. The semantics of a work can also be *Informational* without the arena being *Extended*, as happens in Magnusson's performances with the *Threnoscope* system (see chapter 4.4.3).

The model of this thesis can use a single axis for *Expressive* and *Environmental Semantics* because 'expressive' denotes a focus upon the performer, and 'environmental' denotes the opposite. *Informational* semantics require an individual axis, inclusively because they can reinforce *Expressive* or *Environmental* semantics. For example, a digital 3D world conveys environmental semantics because the graphics are loaded with information about an imaginary world. Also, a close-up of the performer's hands projected on the screen conveys expressive semantics because it draws attention to the performer's gestures.

In my work, the semantics of music are more complex to assess than the semantics of the image and the physical setup. I proposed the term figurative sounds to describe recognisable sounds. Figurative sounds convey what Schaeffer called causal and semantic listening [1966]. My music articulates figurative moments with moments of more abstract nature, which convey "reduced" perception, stripped from conclusive causes and meanings. So how does the perceptual oscillation between figurativism and abstraction contribute to the impression of movement in music? Robert Hatten's definition of musical gesture is inspiring: musical gesture is "any energetic shaping through time that may be interpreted as significant" [2006:142]. The problem is his notion of 'significant': according to him, musical gesture means that "for some interpreter, the gesture will convey information with respect to affect, modality, and/ or communicative meaning" [2006:142]. This understanding constrains the notion of musical gesture to causal and semantic perception. Causes, concepts and meanings are important when we listen to music, but that is solely one aspect of the experience. At points they come to the focus of conscious awareness, and other times they submerge into the perceptual background, as we focus on the experience itself.

Musical gesture can rather be understood as musical motion. Snyder described musical motion as a continuous oscillation between points of high intensity and points of low intensity [2001]. I proposed that intensity derives from the combined effects of automatic and deliberate attention, driven by the interplay of continuities and discontinuities (chapter 3.4.1 and 3.4.2). My adopted musical forms and digital constraints were described using the taxonomy of the thesis, which relates continuities/ discontinuities, intensity, attention and semantics. The semantics of each musical form were bound with the interplay of continuities and discontinuities.

A single, highly figurative sound in an abstract soundscape would create a radical, semantic discontinuity - a point of high intensity. My playing creates points of high intensity, but not at a figurative level. The semantic load of a sonic event changes as the event combines with other events, or as a person changes perceptual resolution. For example, when figurative sounds combine so as to reveal a musical phrase, attention might drift from meanings and causes so as to focus on the qualities of sound organisation itself. My music combines *Informational*, *Expressive* and *Environmental* semantics. The informational load of my figurative sounds can convey expressive or environmental semantics: it can draw attention to the performer, or shift the focus to space. Expressive semantics dominate when the timbre of the zither draws attention to my physical playing, when the music exhibits many discontinuities, or when I address the unexpected in a skilled way. Environmental semantics dominate when sounds of nature recall natural environments, when sonic continuity creates a sense of immersion, or when great sonic complexity leads attention to shift away from detail, so as to embrace the soundscape at lower perceptual resolution.

The parametric model of the thesis can represent sonic constructions by visualising the range of *Sonic Discontinuities*, quantifying *Informational Semantics*, and visualising the proportion between *Expressive* and *Environmental Semantics*. Beyond personal practice, every sonic construction can be described according to these parameters.

6.3 Representing the Practical Work

This section uses the parametric visualisation model to represent my perspective over the practical work, including instrument design and solo performances (Figure 44, please see ahead). I will detail how each aspect of the graphic representation reflects multiple aspects of the creative thinking.

The *Sound Organisation* axis indicates that the interaction with my instrument entails medium *Real-time Effort*, reflecting the principle of sonic complexity: control and unpredictability are thresholded so that effort conveys expression. The model considers effort relative to cognitive processing and musical timing, as well as the skills required to play the instrument/ system. My instrument requires skills to play the zither, and integrate unexpected digital sounds in the musical logics. But the interaction is not very demanding in terms of cognitive processing, and there is immediacy in the construction of musical time:

my direct control over the zither sound makes the instrument extremely responsive, and digital constraints rule out undesired outcomes.

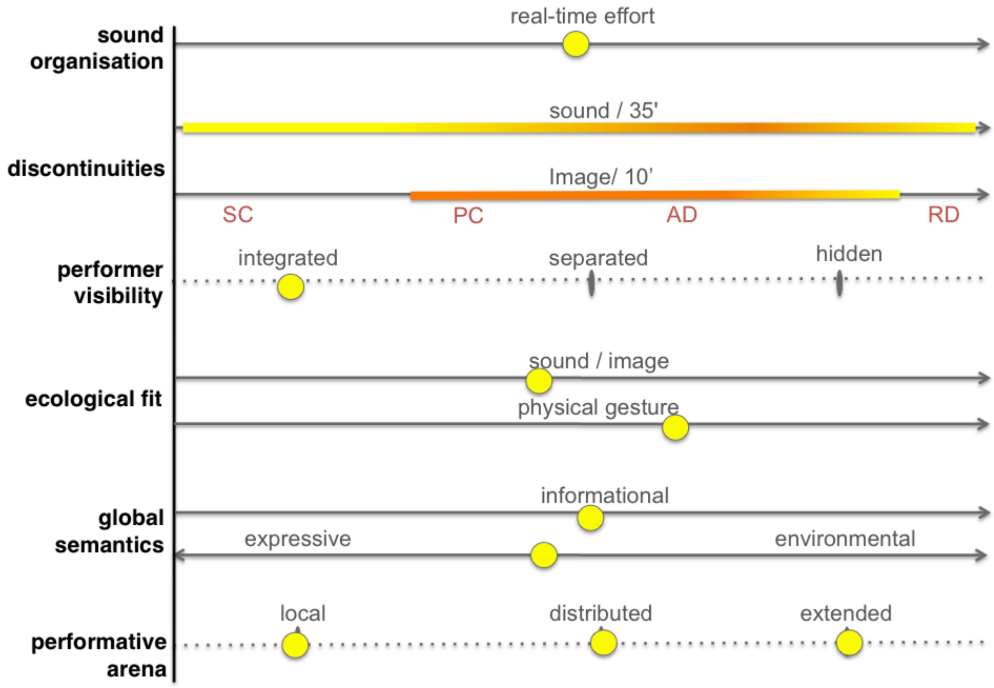


Figure 44: Parametric visualisation of a solo performance from my creative perspective

The two *Discontinuities* axes represent the sonic and visual dynamics. My music combines a few steady continuities (SC), with no intrinsic motion; a few progressive continuities (PC), where successive events display a similar interval of motion; many ambivalent discontinuities (AD), which shift the foreseeable logic of the music without disruption; and occasional radical discontinuities (RD), which are disruptive, prompting automatic attention. The orange gradient on *Sonic Discontinuities* shows that the music is more discontinuous than continuous: the strongest orange is closer to radical discontinuities than to progressive discontinuities. The yellow gradient reaches the left extreme and the right extreme of the axis, showing that the music combines all types of continuities and discontinuities. In my experience, the visual motion solely exhibits ambivalent discontinuities and progressive continuities. Both are constant in the visual dynamics, hence the stripe on *Visual Discontinuities* is mostly orange; the yellow gradient to the right represents the occasional moments where the visual changes become more intense. The axis shows how I assess that the work complies with the principle of visual continuity: the image exhibits no radical discontinuities, which would attract automatic attention and subordinate audition.

Because I don't look at the image when performing, the axes for sonic and visual dynamics reflect different timescales. Whereas the *Sonic Discontinuities* axis indicates the timescale of a performance (about 35'), the *Visual Discontinuities* axis indicates the timescales of studio work and short documentation videos (average 10'). The latter are much shorter, which means that my creative planning is speculative regarding the audience's experience. A disparity seems particularly possible because of the digital "bug" described in chapter 5.5: occasionally, non-numbers produce disproportional terrain elevations. My assessment about their intensity might diverge from the audience's due to the different timescales of our experience, as well as each person's predisposition for attending the visual details. Nevertheless, points of higher visual intensity won't coincide with points of low sonic intensity because the audio detection threshold is high. And the image will always convey Gestalts of invariance, unlike the musical dramaturgy. To some extent, facing this technical incident as creative material is also a relief, because as an artist I am more driven to raise questions than to manipulate the audience. To clarify how perceptual experiences might converge and diverge one could conduct user studies, and define time-based thresholds between ambivalent and radical visual discontinuities. That is beyond the aims of the thesis, but it is good to see how using the model to represent one's creative thinking can lead to new questions.

Returning to the present graphic visualisation, the *Performer Visibility* axis represents my location relatively to the visual projection. The yellow dot on *Integrated* reflects that the image is projected over me, functioning as a stage scene. There is no dot on *Hidden* because I am always visible, and no dot on *Separated* because I never exit the frame of the visual projection.

The two *Ecological Fit* axes indicate medium fit, corresponding to a fungible audio-visual relationship. The dot on *Sound/ Image* represents the fungible audio-visual mapping, as well as the fungible relation between 3D sound (i.e. digital sound spatialisation dependent on visual dynamics) and inverted stereo diffusion. The mapping combines synchronised components, partially related components, and unrelated components. It creates a sense of causation, and confounds the base cause and effect relationships. The same happens with the audio-visual relationship in space. The combination of 3D sound and inverted stereo diffusion creates a sense of causation because the speed at which the sounds move through the speakers equals the virtual camera speed. At the same time, it confounds the cause-effect relationships because the direction of the sound source does not fully correspond to the visible sound emitter in the digital 3D world, i.e. the visible elevation in the foreground terrain (see chapter 5.6). We can hear each sound from two opposite directions, but only one corresponds to the sound emitter on the screen.

The dot on *Physical Gesture* is more to the right than the dot on Sound/ Image, due to the direct correlation between my gestures and the acoustic zither sound. The dot still indicates medium fit because that correlation is not clearly perceivable, for three reasons. First, my visible gestures are often not proportional to the loudness of the zither sound. Second, at times I "disappear" within the visual projection, particularly as I dress in white: my shadow over the projection outlines my physical body, yet my gestures are not clearly perceivable because the body is also a projection surface. And third, the digital sounds are at times undistinguishable from the acoustic sounds, namely when I activate audio sample bank 2, which includes recordings from the zither.

The two axes from *Global Semantics* indicate that the work has *Informational*, *Expressive* and *Environmental* semantic dimensions. The yellow dot on *Informational* is in a middle position, and that results from evaluating several aspects. The fungible audio-visual relationship has informational load, because it explores perceptual cues so as to create a sense of causation; but not much, because the base cause-effect relationships are confounded. The figurative aspect of the sonic construction has informational load; but not much, because one is also led to strip perception from conclusive causes and meanings. The image is more loaded, as it represents an imaginary 3D world.

The yellow dot on the axis for *Expressive* and *Environmental* semantics shows that the expressive dimension of the work is slightly stronger than the environmental: the focus upon the performer is stronger than the focus on space. The expressive dimension of the work derives from a) the audience's sitting position, directed to me; b) my central position relatively to the visual projection; c) the pathos of my playing techniques; and d) the sound source placed next to me on stage (the zither amplifier). The environmental dimension comes from: a) the digital 3D world (an environment per se) and the overall continuity of visual dynamics; b) the size of the projection and my integrated position in front of the screen, in the threshold to the virtual world; c) the environmental aspects of the sonic construction; and d) the sound system placed around the audience.

The *Performative Arena* axis shows a yellow dot on each discrete point, indicating that the work drives attention to shift between the performer, the environment, and imaginary spaces beyond the physical performance space. The arena is *Local* because there is an amplifier on stage, and the music often leads attention to focus on the performer's expressiveness in physical space. The arena is *Distributed* because a double inverted stereo system is placed around the audience, and attention often drifts from the performer, so as to focus upon the environment, or internal states. The arena is *Extended* because there is a big projection of a digital 3D world that morphs with sound; the audio-visual relationship explores perceptual cues, conveying spatial presence in the virtual space beyond the screen; and also, certain digital sounds (sample bank 1) convey a mental transportation to natural environments.

The *Performative Arena* axis provides information that the other parameters do not specify. The dot on *Local* shows that at least sometimes the performer has protagonism over the environment; this specifies *Performer Visibility*, with its dot on *Integrated* – indeed, in Ikeda's *Superimposion* the performers are integrated, but not protagonists. Furthermore, the dot on *Distributed* specifies that a sound system is placed around the audience – indeed, the environmental semantics of the work could derive solely from the qualities of sound and/ or image. Also, the dot on *Extended* indicates that the image has figurative qualities – indeed, bi-dimensional abstract graphics would not extend the arena. The arena could be extended through the sonic construction alone, yet the dot on medium *Ecological Fit* permits disambiguation, as it denotes that the audio-visual relationship applies perceptual cues.

Photo 17: Solo performance at Amersham Arms, London, 2013

When parameterising an audio-visual performance work we must consider every detail, so as to summarise the combination with axes and dots. The model's level of abstraction encourages an analysis of the work as a whole, and that in turn requires an analysis of each individual aspect.

My solo performances are never quite equal, but the graphic parameterisation of my creative perspective should be the same, given the model's level of abstraction. The digital 3D world, the visual dynamics, the audio-visual relationship and the physical setup are similar in every performance with AG#2. Set prior to performance, they concern the creative principles of visual continuity and audio-visual fungibility. During performance I solely focus on the sonic construction; as the creative principle of sonic complexity requires me to address the unexpected resourcefully, I pay full attention to the music. I could use any zither technique combined with any audio sample bank at any time, but decided to adopt specific combinations so as to explore particular musical forms, which have characteristic dynamic and semantic qualities. To some extent, those parameters of the work can change when I play with other musicians. Indeed, the principle of sonic complexity admits any types of dynamics and semantics. The question then is how the parametric model of the thesis can represent this versatility.

6.4 Analysing the Work's Musical Versatility: Four Case Studies

This section aims to clarify how my instrument affords an adaptation to different musical situations, using a same set of pre-recorded sounds, mappings and configurations. Simultaneously, the section aims to clarify if, and how, the parametric model can represent how I seek to enable, and put limits to the versatility of my audio-visual performance work.

The previous section of this chapter represented my solo performances, and most parameters will remain the same regardless of whether I'm playing with other musicians or not. My interaction with the instrument maintains a same level of *Real-Time Effort,* and the work combines the three types of *Performative Arena*. The projected image is also similar, exhibiting a same range of *Visual Discontinuities*. In the context of this investigation we can simply "scale up" my apprehensions: the intensity of visual changes might change according to the timescale of one's perceptual experience, but the amount of disproportional terrain elevations derived from non-numbers does not depend on whether I play solo or in collaboration. Furthermore, the fungible audio-visual relationship is assured; there is a medium level of *Ecological Fit* even when the other musician's physical gesture fits highly with their sonic emission. There are two reasons for this. First, the piezo microphone attached to the zither also captures the sound of the room, thus at times the software reacts upon the other musician's sounds, rather than mine. Second, their physical gestures are not clearly distinguishable because the image of the 3D world is projected upon us both. Consequently, the overall audio-visual relationship combines components that convey causation and components that do not, exhibiting complexity and inconsistency. Given all these constant parameters in the work, the question of versatility is narrowed to how the collaboration with another musician affects the dynamics of the sonic construction, and semantics as a whole. In the parametric model, that should solely reflect in the axes for *Sonic Discontinuities* and *Semantics*.

This section presents a set of case studies in which I play with different musicians, each with particular aesthetics and instruments. I use the Arpeggio-Detuning software, i.e. the audio component of AG#2. In each case study, I will use my taxonomy of continuities and discontinuities to analyse the music. Subsequently, I will use the model to estimate how the duo would work in the context of a performance with my audio-visual instrument, including the visual component of AG#2. In the model, the axis for *Sonic Discontinuities* concerns the music alone, yet the axes for *Semantics* concern the work as a whole, including image, audio-visual relationship and physical setup. To estimate how the music would affect the overall semantics of an audio-visual performance, I will add its semantics to the semantics of the image and the physical setup, which are the same in solo performances and collaborative performances.

6.4.1 Question, Procedure, and a Distinction Between Expressive and Structural Changes in Music

My work explores musical forms developed through the combination of specific audio materials and zither playing techniques. These musical forms are versatile, and I can use them to create different dramaturgies. I can also use the same combinations of digital sounds and zither techniques when I play with other musicians. Performing with another musician is always very different from performing alone, yet the parametric model of the thesis should not visualise all types of differences, given its level of abstraction. If another musician's playing is consonant with mine, the graphic parameterisation of our duo should be equal to the one of my solo performances. Otherwise, it should be not. Defining the threshold between consonant and dissonant transformations in music should clarify the model's level of abstraction.

To define this threshold, we should not rely on any distinction between composition and instrument. At first sight, the notion of instrument seems to imply greater versatility than the notion of composition. But for example, a didgeridoo (wind instrument developed by Indigenous Australians) is not really versatile. And, graphic scores can leave room for musicians to choose their own material, or the sequence in which they use that material. Pressing's notions of hierarchical and dynamic complexity [1987] seem more useful to define a threshold between consonance and dissonance. Dynamic or adaptive complexity refers to a rich range of behaviours over time, an adaptation to unpredictable conditions, or a monitoring of results in relation to a reference source. And hierarchical complexity structures converging and diverging sonic information as a cohesive whole (see chapter 3.3.2). We can interpret that hierarchical complexity exists at a low level in the music – at a structural level. And we can say that dynamic complexity exists at a higher level – at an expressive level. Our parametric model should solely visualise differences at a structural level, revealing if the collaborative music is dissonant from the spirit of my solo music. The thesis' taxonomy of continuities and discontinuities can be used to describe the music, and distinguish an expressive and a structural level of transformations.

Another musician affects my music at a high level when his or her playing shifts my phrasings without disruption. The difference between the collaborative music and my solo music substantiates in ambivalent discontinuities. This type of discontinuities shifts the foreseeable logic of the music, yet not disruptively. The other musician's playing can be very present, but it draws from the characteristic qualities of my musical forms. That defines what I call an expressive type of transformation: the collaborative music can differ from my solo music in many ways, but the proportion of continuities/ discontinuities and the overall semantics are similar. Hence, the parametric model will not visualise the difference.

The collaboration with another musician transforms my music at a structural level when it twists the spirit of my musical forms. Structural changes have an impact upon what I have called endogenous continuity (chapter 3.4.2). Endogenous continuity enables us to sense cohesion in the music, as perception ties sonic events together, relating them at a higher level of information processing. Transformations at this global level must reflect in a different proportion of continuities/ discontinuities, and a semantic shift. Hence, the parametric model will visualise the difference.

The following case studies aim at gleaning if, and how, the collaboration with other musicians affects my music at an expressive and/ or a structural level. In each occasion, I used my three audio sample banks (common to the Arpeggio-Detuning and AG#2) in the same order: bank #1, #2 and #3. The other musician was asked to play however they thought worked best, without any further explanation about this investigation, or its methods. In each case study, my analysis is based on an audio recording. The recordings are at http://adrianasa.planetaclix.pt/research/ArpeggioDetuning.htm#music

6.4.2 Case Study #1: Collaboration with Joanne Cannon

Joanne Cannon came for a residency at my studio in Lisbon, during which we developed a collaboration initiated at Goldsmiths. She brought a bassoon and a custom instrument called *serpent*, which she describes as a digital wind controller instrument. Joanne's serpent combines acoustic and digital elements like my instrument, but her approach to interaction design is very different. The controllers attached to the acoustic body of the serpent enable full control over the digital sound output. Moreover, Joanne sees the acoustic sound as a controller, and its intrinsic qualities are less significant. The serpent has been developed with her husband Stuart Favilla, along the musical idiom of their Bent Leather Band[69] duo. Jon Rose described that idiom in a very Australian-like way: "The music stopped and started, fumed and erupted, turned off at obscure angles like a reconstituted 1960's Holden fitted with two steering wheels in the hands of a couple of jackaroos (Jon Rose, 2002).

Photo 18: Joanne Cannon playing the Serpent

The collaboration with Joanne was challenging at first; and then rewarding, as we managed to explore the tension between our different sound qualities and playing manners as creative material. As we progressively got acquainted with each other's sensibilities, techniques and concerns, we also developed unique musical shapes. Prior to the recording analysed in this thesis[70], we discussed a musical structure where at times both play acoustic and digital, and other times, one of us drops the digital sounds to play solely

[69] http://www.bentleather.com/
[70] Audio recording (38 min) at http://www.cityarts.com/audio/sa&cannon.mp3

acoustic. Our work developed into a piece, which was performed in Lisbon and Porto. I will constrain my analysis of the recording to the sections where I combine acoustic and digital sounds, because this case study respects to how those musical forms may change in collaboration with other musicians. Yet, a short description of the whole is useful to situate these sections with respect to their original context.

The recording starts with Joanne on the serpent while I play the bow, acoustic. It takes time before I introduce the sounds of water and thunder, with sample bank 1; first they combine with bow, then with dribbling. Joanne switches to the bassoon while I play bottleneck and pick. I deactivate the digital sounds so as to play solely acoustic; then I activate bank 2, which includes short string instrument sounds. Subsequently, Joanne returns to the serpent as I play bow and bottleneck, with the piano notes and digital timbres from bank 3.

Photo 19: Performance with Joanne Cannon at Trem Azul, Lisbon, 2013

The electronic sounds from the serpent are incisive and funny, with big amplitude variations. They tend to create radical discontinuities, which prompt automatic attention. In contrast, my musical forms with sample banks 1 and 3 are dense and textural. They create overall continuity, with a wealth of ambivalent discontinuities, inviting for deliberate attention. These are cancelled if attention is constantly being attracted by radical discontinuities. The creative challenge was to create an interesting attention dynamics without Joanne's serpent becoming simply a protagonist. It was acceptable to cancel the environmental dimension of my musical forms, but it would not make sense to shift my playing to the background as accompaniment.

Joanne played slower than usual, giving space to subtle mutations in the music, yet maintaining her characteristic energy and spirit. I adapted my forms to engage with that spirit, and created more radical discontinuities than usual. At times our musical phrases unfold in independent directions, and other times they converge in unison. Overall, points of unison happen at a level of detail, inviting attention to focus on ambivalent discontinuities. One of us may pick up on a phrase from the other, and complete it in an unexpected way, as if the music were an exquisite corpse game. This creates a paradoxical sense of completion. On the one hand, completion is characteristic of musical closure [Snyder 2001]: psychophysical expectations are fulfilled. On the other, our musical idioms report to different musical traditions, which leads completion to happen in unexpected ways, counteracting expectations.

Our musical shapes were easier to balance in the middle section of the music, when Joanne played the acoustic bassoon. Its timbre is much softer than the serpent's; it brings a melodic bass-body to the music. Also, in this section my playing is more expressive than dense and environmental. The zither sound is incisive when I use the metal pick; and the short digital sounds tend to produce discontinuities, attracting automatic attention.

The collaboration with Joanne shows that my musical forms can change at both an expressive and a structural level. Throughout the whole audio recording, Joanne's playing affects those forms at an expressive level, as we pick up on each other's musical phrases. Her bassoon playing has an expressive impact without twisting their original spirit. Conversely, her serpent playing has both expressive and structural impact. The structural change is due to the great amount of discontinuities and to the playful timbres and timings, which counteract the environmental dimension of my musical forms with bank #1 (water and thunder sounds) and bank #3 (long digital timbres).

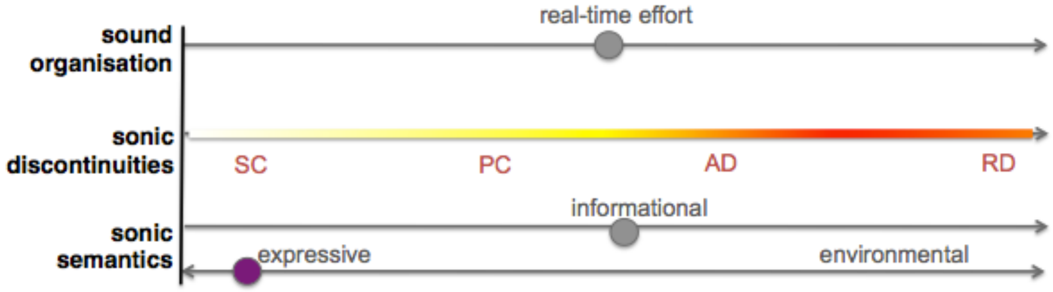


Figure 45: Parametric visualisation of the musical collaboration with Joanne Cannon

Given this structural change, the graphic parameterisation of our collaborative music differs from the one of my solo music. Figure 45 shows the difference. Whilst the gray dots correspond to aspects that remain similar, the coloured dots correspond to changes. The yellow/red gradient on *Sonic Discontinuities* shows more radical discontinuities, and less continuities. And the purple dot on *Semantics* shows that the semantics of the music is now much more expressive than environmental.

We can also use the model to visualise how this duo would work with the visual component of AG#2. I have explained that we can do this because the parameterisations of visual dynamics, audio-visual relationship and physical setup are equal whether I perform solo or collaboratively. In particular, I explained that the software at times reacts upon the other musician, assuring an overall fungible audio-visual relationship. But regardless, Joanne's serpent conveys a medium level of audio-visual fit in itself (Figure 46), for two reasons. Firstly, because the semantics of her acoustic controller contrast with the semantics her electronic sounds. And secondly, because the acoustic controller is barely audible, thus we cannot clearly perceive a dynamic musical relation between acoustic and digital sounds.

Figure 46: Ecological fit intrinsic to Joanne Cannon's Serpent

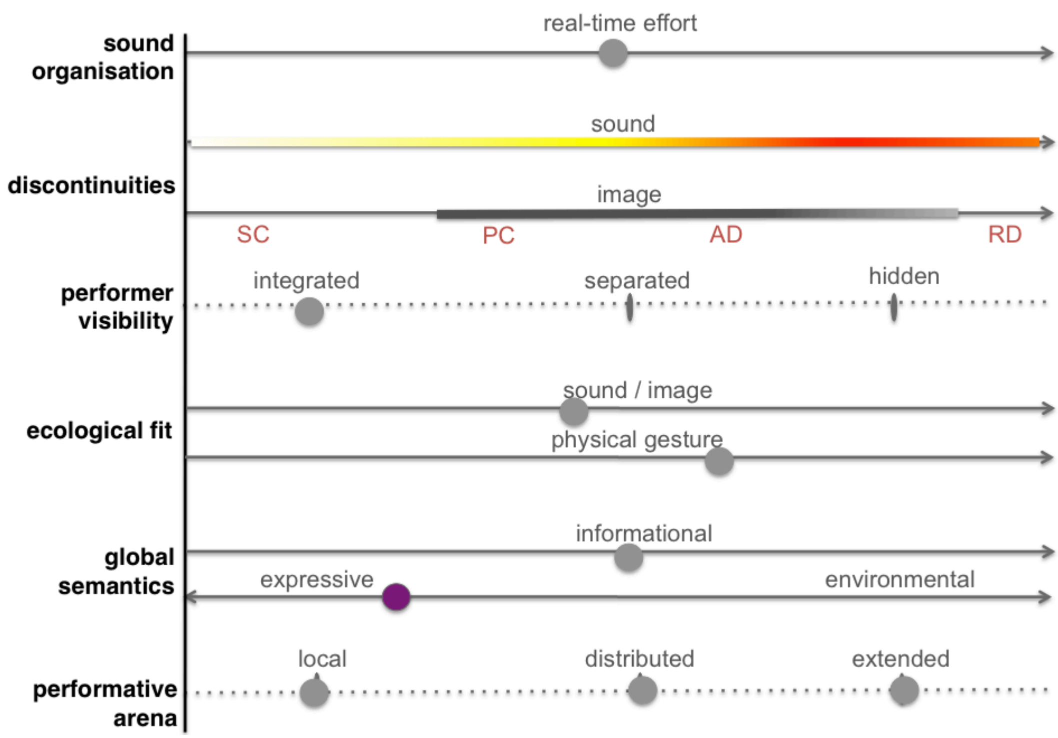


Figure 47: Parametric visualisation of an hypothetic audio-visual performance with Joanne Cannon

Figure 47 represents a hypothetic audio-visual performance with Joanne. It shows the transformation in sonic dynamics, as in Figure 45. It also shows that the overall semantics of the work are more expressive than environmental. However, the purple dot is less to the left than in Figure 45, because the digital 3D world of AG#2 and the speakers around the audience create highly environmental semantics.

6.4.3 Case Study #2: Collaboration with Helena Espvall

Helena Espvall[71] is a Swedish cellist (and guitarist), who lives in Lisbon since 2012. She played in rock bands and has formal training in music, but never wished to be a classical cellist. In her words: "My background as an electric guitarist made it natural for me to drag the cello into a more unconventional realm, and use effect pedals and an improvised approach. A couple of years in an Arabic music ensemble have also deeply affected the way I play the cello" [private email, October 2013].

Helena came for a studio recording[72] and then kindly answered a few questions posed by email. Each musician from these four case studies was asked for the same, but to my surprise, Helena was the only one who responded to my interview. I interpreted that my questions were difficult because they required descriptions of music. Music can be described in very evocative and imagetic ways, as we will see with Helena. I wanted to see if the other musician's wording could be related with the three semantic dimensions of our parametric model – expressive, environmental and informational. I also wanted to see if their wording would provide cues with respect to my distinction between expressive and structural transformations in music. In summary, I wanted to see if my music analysis method would be useful to clarify the words chosen to describe music.

Question #1 aimed to provide cues about whether the musician was interested in exploring a tension between personal modes of musical expression, or if she preferred to adapt her techniques so as to explore a common musical identity:

A - Did you change/ adapt any particular aspect of your techniques in our collaboration?
H - *Yes, I changed the tuning of the cello to be closer to that of the zither*.

Question #2 aimed at an overall semantic description of the music:

A- How would you describe the music in the recording?

[71] http://en.wikipedia.org/wiki/Helena_Espvall
[72] Audio recording (20 min) at http://www.cityarts.com/audio/sa&espvall.mp3

H - *Moody, emotional, cinematic, sometimes eerie or ominous*.

The terms ‘moody’ and ‘emotional’ reflect human-scaled, personal states; they imply a focus upon the performer, which corresponds to expressive semantics. In contrast, the terms ‘eerie’ and ‘ominous’ connote some sort of “otherness”. Helena’s terminology recalls horror movies, i.e. frightening environments. We can say that the music has an informational load that conveys environmental semantics.

Question #3 aimed at a description of my musical forms. I wanted to see if it converged or diverged from the Helena’s description of our music. A divergence would possibly indicate a transformation at a structural level.

A - My software includes three sets of digital sounds. I used these sets sequentially, with different zither techniques. With set 1, the zither was dribbled or played with the bow. As an input to the software it activated sounds of bass guitar, ocean waves, water drops, thunder (synthesizer) and wind (synthesizer). With set 2, the zither was played with hands, bottlenecks and pick. The digital sounds were from dobro, bass guitar, and zither (played with bottleneck and metal pick). And with set 3, the zither was played with bow and bottleneck, activating piano notes and digital timbres. How would you describe my musical forms in these three sections?

H - *In the first section, it appears to me like the nature sounds are reflected in the music, like the zither sound here bears a resemblance of water, flowing waves. In the second section, maybe due to the bottleneck slide playing, the zither sounds somewhat drunk and drowsy/wobbly, a dreamlike quality of slowly falling. In the third section, the high pitched bowing of the zither at times makes me think of a medieval fiddle or earthy droney folk music, a nice contrast to your haunting otherworldly samples. In general, the old detuned zither strings seem to evoke something decrepit but beautiful, the ghosts of an abandoned building… and the sample sounds add a richer texture, more flavours as well as unpredictable juxtapositions, an aleatoric aspect that i find compelling.*

Helena’s writing is very imagetic. Her description of my musical forms with sample bank 1 connotes environmental semantics – the zither dynamics follow water dynamics, hence the focus is not at human scale. Her description of my forms with bank 2 impersonates the zither with human characteristics, implying a focus upon the performer. And in her description of my forms with bank 3, the expression ‘haunting otherworldly’ matches her previous terms ‘eerie’ and ‘ominous’, denoting environmental semantics. In summary, her description of my musical forms seems to converge with her overall description of our collaborative music.

Questions #4, #5 and #6 aimed at revealing if the other musician felt a tension between our personal modes of musical expression, or not.

A - What did you think of the music in terms of musical flow?

H -*To me, the flow of our music seems very natural and effortless, like breathing.*

A - Can you explain your musical decisions in each section?

H - *My playing is very intuitive, and there's not a lot of conscious decision-making going on in my case. Rather, I tend to enter a near trance-like state and try to stay out of the way of the music, to let it find its own path without interfering too much.*

A - How did the combination of our musical languages work for you?

H - *It seems to me like our playing styles blend quite well, and I'm personally very fond of the outcome of this recording, which i find varied and evocative.*

A - Were your musical forms in consonance mine, or did you intend for a distinct position, or was it sometimes one and other times the other?

H -*Sometimes one, sometimes the other. Sometimes we seem to mimic each other, and weave in and out of each others' textures in a twin-like way, other times we are doing opposing things, my nervous energy contrasting with your calm, to best complement each other, whatever the music seems to call for.*

[private email, October 2013]

Photo 20: Recording with Helena Espvall at ScratchBuilt Studios, Lisbon, 2014

Helena interpreted that ‘playing in consonance’ means mimicking each other: “sometimes we seem to mimic each other (…) other times we are doing opposing things”. The term ‘opposition’ can benefit from disambiguation, because I felt that our playing modes were always in agreement. Helena’s wording also expresses this agreement: “very natural and effortless, like breathing”, plus “I try to stay out of the way of the music”. As Helena wrote, sometimes we play at a different pace. But there is never a sense of counteraction, contrary to what happens in the recording with Joanne. For a start, the timbres of the zither and the cello fit naturally - both are string instruments. Helena also changed the tuning of the cello to be closer to that of the zither. Overall, the collaborative music seems in consonance with my solo work, and my taxonomy of continuities and discontinuities should help to explain this convergence.

In section 1 of the recording I dribble the zither, activating sounds of bass guitar, ocean waves, water drops, thunder and wind. Helena’s bowing adds textures and tonalities without ever producing radical discontinuities. This is characteristic of environmental semantics. Helena captures movements of timing and tonality that are implicit in my forms, and makes them explicit by developing those movements until they reach musical closure (see [Snyder 2001]). Our phrasings complement each other so as to create Gestalts of good continuation (see [Wertheimer 1938]). At times Helena plays fast, increasing the density and complexity of my musical forms. As perception lowers perceptual resolution to embrace the whole, ambivalent discontinuities decrease in intensity. This is characteristic of environmental semantics as well. Occasionally, we let density faint away, which invites attention to focus on detail. We can say that when I play solo, the zither/samples combination #1 originates similar attention dynamics and semantics.

In section 2 I play the zither with hands, bottlenecks and pick, activating audio samples from dobro, bass guitar, and zither (played with bottleneck and pick); the samples are tonal and relatively short. There is great variety in Helena’s playing throughout this section. At times she uses the bow, and other times she plucks the strings in a jazz-like manner. I create discontinuities and intervals, increasing the intensity of individual sounds and phrases. She creates new levels of dynamic complexity, incorporating my sounds in her phrasings. At times perceptual resolution may decrease to embrace the whole, but never for a long time, because one’s attention is constantly being attracted by new discontinuities. We can say that when I play solo, the zither/samples combination #2 affects attention dynamics and semantics in similar ways: this section of the music is expressive, leading to a focus upon the performer.

In section 3 I play the zither with bow and bottleneck, activating piano notes and digital timbres. At first Helena plucks the strings creating a stream of rapid phrasings. Then she uses the bow to create circular motion, repeating a phrase consistently, with harmonic deviations. Her speed contrasts with my calmness, and her phrasing feels inlaid in the mass of sound, as we keep forming new tonal relationships. While the sonic density creates a space of overall continuity, conscious awareness is invited to focus upon ambivalent discontinuities, which makes chromaticisms and harmonics intense. This is characteristic of environmental semantics. Again, we can say that the semantics of the music and attention dynamics are similar when I play solo.

Helena's impact upon my musical forms is very different from Joanne's. Whereas Joanne counteracts the environmental dimension of my musical forms with sample bank 1 and 2, Helena's "nervous playing" (sic) lets perception focus on ambivalent discontinuities, which are important in those forms. Whereas Joanne takes my phrasings in unexpected directions, Helena goes in consonant directions. My musical analysis method seems useful to clarify the evocative words of a musician. Indeed, Helena affects my musical forms at an expressive level, but there is no structural transformation. The parametric model cannot visualise the difference between our collaborative music and my solo music, given its abstraction level. Accordingly, the parameters in Figure 48 appear gray/ black.

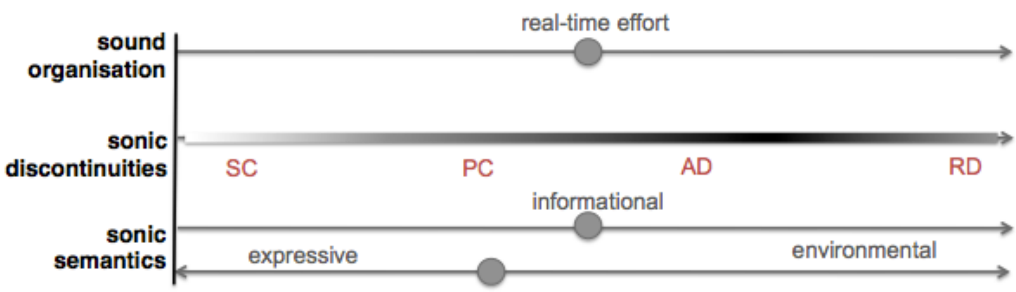


Figure 48: Parametric visualisation of the musical collaboration with Helena Espvall

If we use the model to analyse her solo work, the *Ecological Fit* looks high because the visible cello fits with the cello sound (Figure 49). This will not happen in a performance with my audio-visual instrument. First, sometimes AG#2 reacts upon her cello, rather than upon my zither. Second, the moving image is projected over her, just like over myself; thus her gestures are at times not clearly discernible. And third, the base cause-effects are even more confounded because the sonic construction combines two acoustic string instruments, plus digital recordings of string instruments.

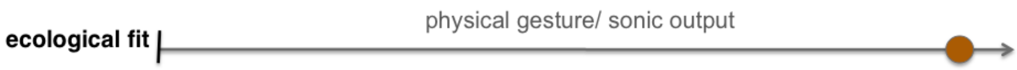


Figure 49: Ecological fit intrinsic to Helana Espvall's cello playing

Figure 50 shows the graphic parameterisation of a hypothetic audio-visual performance with Helena. The stripes and dots are gray/black because it equals the one of my solo performances. The dot on expressive/ environmental semantics is more to the right than when we solely consider the music (Figure 48), because the image increases the environmental dimension of the work.

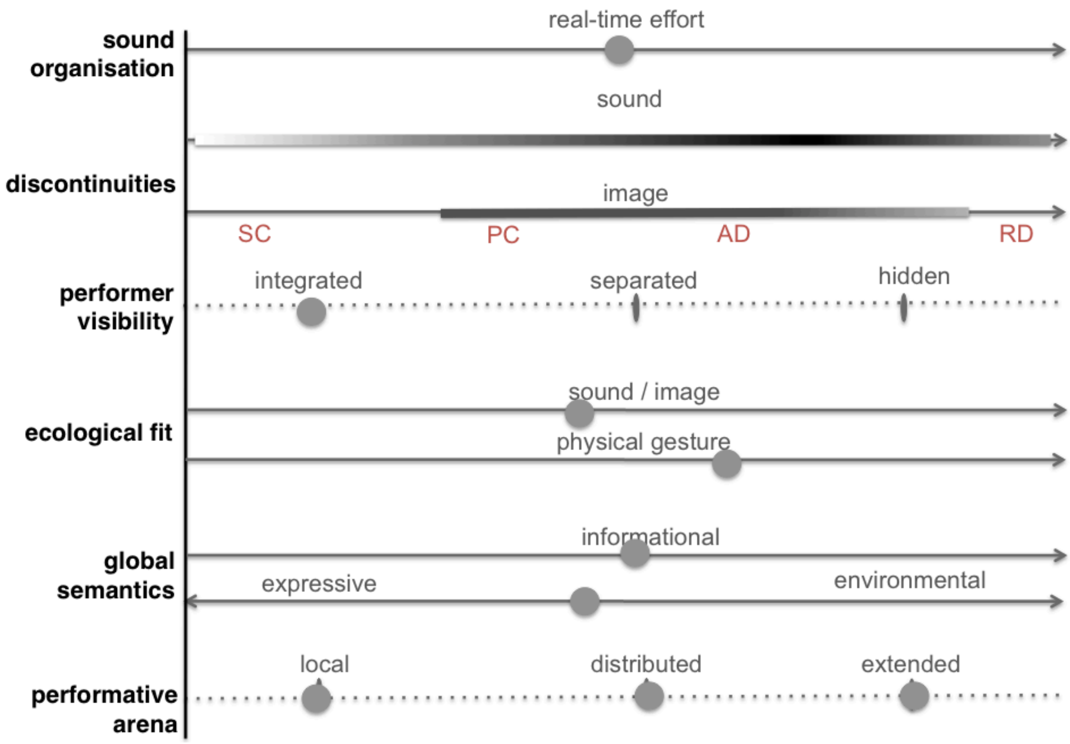


Figure 50: Parametric visualisation of an hypothetic audio-visual performance with Helena Espvall

6.4.4 Case Study #3: Collaboration with Tó Trips

Tó Trips[73] is a well-known guitarist in the Portuguese rock scene; in the last twenty years, he has been a member of bands like *Santa Maria Gasolina Em Teu Ventre*, *Lulu Blind* and *Dead Combo*. His music has been labelled rock, folk, blues, or fado – it draws upon all

[73] Video of a solo concert by Tó Trips: https://www.youtube.com/watch?v=uCP4QeRUTKk

those styles without confining to any in particular. The first time we played together, prior to this investigation, I felt surprised with his amazing aptitude to explore new musical territories; he changed his tuning according to the zither, and dispensed with metrics very easily. Together with John Klima, we formed a trio and made a record called *timespine*[74] in which we play zither, dobro and bass. Prior to this case study, I had never used software to play with Tó; the recording analysed in this thesis[75] documents the first time it happened.

Photo 21: Performance with Tó Trips at Trem Azul, Lisbon, 2013

In this recording, Tó plays acoustic guitar. He plays with many silences, creating melodic punctuation, with no radical discontinuities. He develops chromatic details, creating tonal accentuations that convey Gestalts of good continuation.

In the first section of the recording I play the zither with hands, combined with the water and thunder sounds from sample bank 1. I play very differently from the way I did with Helena or Joanne: now the zither does not submerge and emerge from water sounds. Instead, I play with large intervals and sparse chords, like Tó. Musical phrases begun with acoustic sounds are completed with sounds from nature, and vice-versa. The acoustic sounds gain

[74] timespine record: http://www.shhpuma.com/timespine/
[75] Audio recording (17 min) at http://www.cityarts.com/audio/sa&trips.mp3

greater presence, but in consonance with the dynamics of digital sounds, emphasising the environmental dimension of the music. Our space-giving intervals bring a dramatic flavour to the natural sounds, so that human gestures gain environmental scale. Perception is invited to focus on ambivalent discontinuities, as happens when I play solo.

Throughout section 2 and 3, Tó continues picking up chromatic variations so as to unfold musical phrases, in ways that create Gestalts of good continuation. In section 2 I play with bottleneck and pick on the zither, combined with the short string instrument sounds from sample bank 2. Tó and I pick up on each other's tonalities, creating a circular phrasing. In section 3 I play bow on the zither, combined with the piano notes and digital timbres from sample bank 3. My continuities contrast with Tó's melodic phrases. That could shift the nuances of my playing to the background, because overall continuity leads attention to focus on other, more discontinuous streams of sound. However, Tó's intervals leave space for attention to focus on ambivalent discontinuities. In both sections of the recording, one's attentional movements are similar to when play solo. Accordingly, the semantics of the music are more expressive in section 2, and more environmental in section 3.

Tó plays in consonance with the spirit of my musical forms, like Helena. Both musicians pick certain chromatic details, creating tonal accentuations that convey Gestalts of good continuation. Both engage with the environmental semantics of my musical forms with bank 1 and 3. But interestingly, their playing agrees with my musical forms in very different ways. With Helena, the music is orchestral, unfolding in complex ways, with many layers. Musical closure happens at a textural level. Conversely, Tó plays with many silences, creating melodic punctuation. The musical phrasing unfolds at much slower pace; musical closure happens at a higher hierarchical level in information processing.

This means that the nuances and chromaticisms of my musical forms can convey different musical logics. The music with Tó is quite different from the music with Helena, but both are consonant with my solo music. Hence, the graphic parameterisation is the same (Figure 51). The audio-fit of Tó's guitar playing would be high if he was playing alone (Figure 52), as happens with Helena's cello playing. And again, that would not be the case in a performance with my audio-visual instrument, because: sometimes AG#2 reacts upon the other musician (the piezo attached to the zither captures room sound); the image of the digital world is projected over us (physical gestures are often not clearly perceivable); zither sounds, guitar sounds and digital sounds are often indistinguishable. Hence, a hypothetic audio-visual performance with Tó is parameterised like the one of my solo performances (Figure 53).

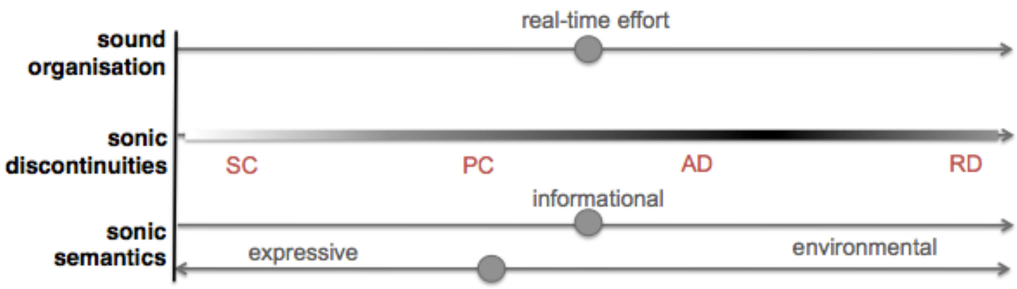


Figure 51: Parametric visualisation of the musical collaboration with Tó Trips

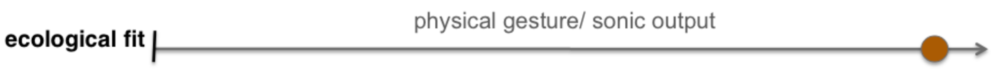


Figure 52: Ecological fit intrinsic to Tó Trips' guitar playing

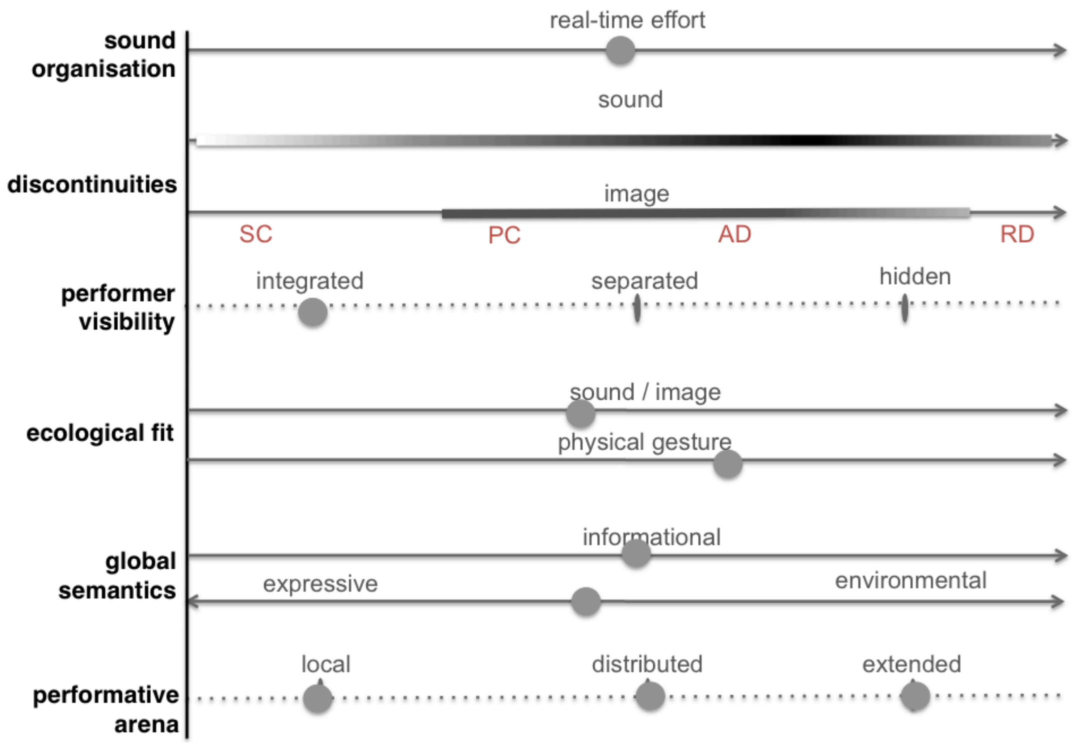


Figure 53: Parametric visualisation of an hypothetic audio-visual performance with Tó Trips

6.4.5 Case Study #4: Collaboration with Aki Onda

Aki Onda[76] is an electronic musician, composer, and visual artist. He is particularly known for his *Cassette Memories* project – works compiled from a *sound diary* of field-recordings, which he collected over a span of two decades. Aki's musical instrument of choice is the cassette Walkman. He captures field recordings with the Walkman, and manipulates multiple Walkmans with electronics. Furthermore, he requires vintage amplifiers to play his music.

Photos 22 - 24: Recording with Aki Onda at ScratchBuilt Studios, Lisbon, 2014

[76] http://www.akionda.net

In our studio recording[77], Aki creates continuous layers of sound, varying their density and loudness, with no disruptive changes. His sounds have often a machine-like, vintage timbre. Aki modulates his sonic emissions in a progressive manner; one perceives them to change, but one cannot really apprehend when the changes occur. As attention is attracted to my quicker phrasings, his progressive continuities tend to become steady continuities.

In the first section of the recording, my zither dribbling submerges and emerges from water, thunder, and long bass sounds. Aki's sound emerges and submerges in that musical form as if he was mirroring my motion, but at a much slower pace – at a different time scale. He starts with a thin, unobtrusive layer of noise, which tends to be ignored, as happens with the noise of a guitar amplifier when the musician starts playing. His sonic amplitude and density increase gradually, transforming the noise into a tonal mass of sound. My playing is progressively absorbed, becoming an almost undistinguishable texture. The density of Aki's drone starts decreasing, slowly, until the emission becomes a layer of 'noise' again. Now we are aware of the 'noise', contrary to what happened in the beginning: we were driven to realise its musical presence and meaning.

I start section 2, combining the short string instrument sounds from sample bank 2 with bottleneck and pick on the zither. Aki plays a continuous bass sound, creating progressive continuity as the sound increases in density. In spite of this density, my phrasings are clear because his bass frequencies do not cancel my mid-range frequencies. At some point he starts creating harmonics, introducing another sort of progressive continuity, from a lower pitch to a higher pitch. At first, the pathos of my playing shifts Aki's sonic emission to the perceptual background. But Aki brings it forward with small "eruptions" of voice, which lead us to think of a recording, or maybe a radio.

These eruptions create ambivalent discontinuities, which do not attract automatic attention per se. However, they have a semantic quality that provokes a perceptual shift: now we wonder about where and how the sound was recorded, and through which device it is being played back. We focus on the informational load of sounds. At points the zither sounds seem equally played back through a vintage machine, rather than created in real-time. In other words, Aki's sounds convey causal listening [Schaeffer 1966, Chion 1994] and that affects the semantics of my sounds as well. Attention drifts from the human act of playing, so as to focus on how our musical systems enable the creation of an environmental soundscape. When Aki's sonic eruptions become rhythmic through mechanical repetition,

[77] Audio recording (23 min) at http://www.cityarts.com/audio/sa&onda.mp3

the music looses the semantic load of a recording. At some point I stop playing, inviting attention to focus upon Aki's subtle modulations of continuity.

Section 3 starts with my bowing emerging from Aki's bass drones, which become rhythmical as I start soloing on top. His rhythmic, steady continuity creates a scenario where my phrasing unfolds; perception segregates my stream of discontinuities. Later we enter unison, when I activate sample bank 3. My long audio samples create a dense soundscape, oscillating in a similar frequency range than Aki's sounds. Aki's steady rhythm shifts to the foreground whenever density faints away. My short piano samples punctuate the soundscape at irregular intervals. The music unfolds in cycles, from regularity to irregularity and back.

Aki does not interact with my musical forms in terms of expressive phrasing: our collaborative music is very different from my solo music, but the transformation does not concern the logics of musical closure. Aki's musical idiom is extremely different from Joanne's idiom, but it equally has a structural impact upon my musical forms. His vintage devices and pre-recorded sound materials prompt causal listening, reinforcing the informational dimension of the music. Attention shifts from the human act of playing, so as to focus on how mechanical functions modulate an environmental soundscape; and that weakens the expressive semantics of my musical forms with bank 2.

Given Aki's structural impact upon my musical forms, the graphic parameterisation of the music differs from the one of my solo work. Figure 54 shows the difference. The gray dot corresponds to that which remains equal, and colours represent changes. The yellow/ red gradient on *Sonic Discontinuities* shows more continuities than in my solo music, and no radical discontinuities. Also, the blue dots on *Semantics* indicate stronger informational and environmental semantics.

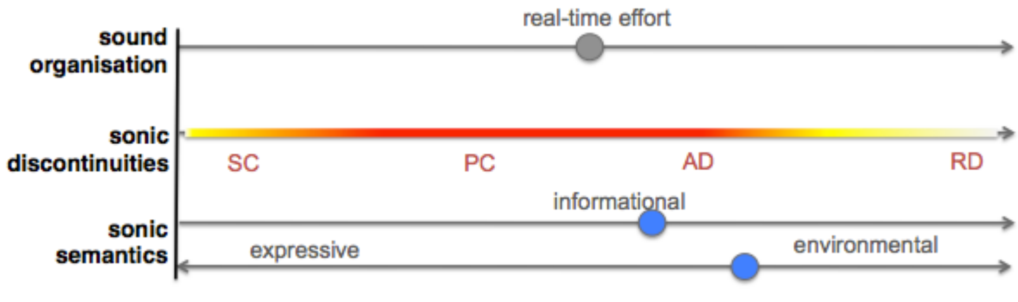


Figure 54: Parametric visualisation of the musical collaboration with Aki Onda

We can use the model to visualise how this musical collaboration would work with the visual component of AG#2. If a regular light bulb were placed next to Aki, one would see his physical gestures. Gestures and sonic emissions would not be consistently synchronised, but if standing close enough one would get the sense of understanding cause-effect relationships, due to his well-known analogue devices. Figure 55 represents the audio-visual fit of Aki playing solo: it is higher than lower.

Figure 55: Ecological fit intrinsic to Aki Onda playing his electronic devices

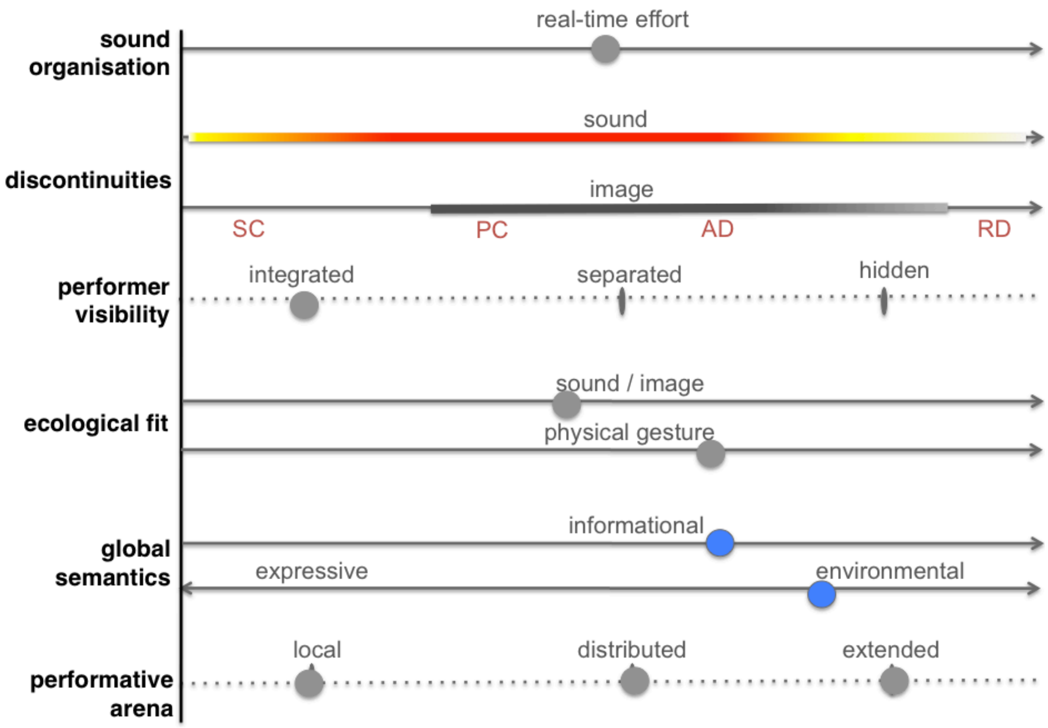


Figure 56: Parametric visualisation of an hypothetic audio-visual performance with Aki Onda

Again, the audio-visual fit intrinsic to Aki's playing would not affect a performance with my audio-visual instrument, because the image would be projected over us both, and because AG#2 at times reacts upon the other musician's sounds as well. Figure 56 shows which aspects of the work would remain similar (gray) and which would change significantly (coloured), considering the parametric model's level of abstraction. The *Sonic Discontinuities* axis shows a change, which is equal to the change represented in Figure

54. The axes for *Semantics* look different from Figure 54: the blue dots on *Informational* and *Environmental* are more to the right. That is because the fungible mapping and the image also have informational load, and the informational load of the image also reinforces the environmental semantics of the work.

6.4.6 Representing Versatility

The instrument developed in this research allows for a wide range of musical explorations. The adopted combinations of zither techniques and digital audio banks convey characteristic types of musical motion when I play solo; yet when I play with other musicians my musical forms can adapt at both expressive and structural levels, without imposing a particular type of musical motion, nor shifting to the background as accompaniment. Recapitulating, the sonic construction changes at the expressive level when the other musician engages with the semantics and dynamics of my musical forms; the difference between my solo sonic constructions and the collaborative sonic constructions is certainly notable when we listen to the music, but the parametric model of the thesis does not visualise this type of difference. The model solely visualises structural differences, which imply a different proportion of sonic continuities and discontinuities, and a different proportion between the semantic dimensions of the work.

The parametric model is useful to estimate how a musical collaboration would affect my audio-visual performance work as a whole. The graphic in figure 57 represents the four previous case studies at once, so as to visualise the work's versatility (please see next page). The gray stripe on visual discontinuities and the gray dots show that certain aspects of the work are not versatile: the visual dynamics, the ecological fit, the physical setup, and the performative arena. The coloured stripes and dots show that the proportion of sonic discontinuities and the semantics of the work can change to a certain extent. Case studies #1 and #4 illustrate the limits of this versatility.

The yellow stripe with a purple gradient and the purple dot represent study #1, with Joanne Cannon. Our collaborative music differs from my solo music at both an expressive and a structural level. We picked up subtle chromaticisms and interacted at expressive level. Also, she was agile in taking my musical phrasings in unexpected directions, in disrupting their logics and pace. Accordingly, the purple/ yellow stripe shows a greater concentration of radical sonic discontinuities. This greater amount of discontinuities and the humorous quality of her electronic timbres and phrasings would alter the semantics of my audio-visual performance work. The digital 3D environment from AG#2 would still convey an environmental semantic dimension, but the expressive dimension would be much stronger. The purple dot shows a corresponding semantic shift.

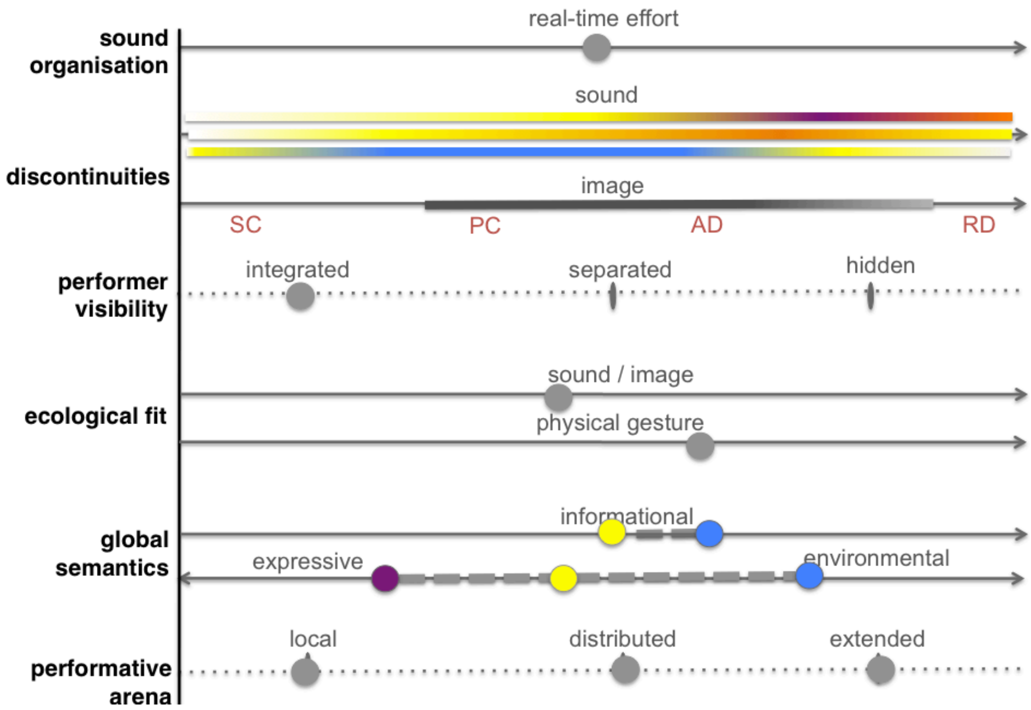


Figure 57: Parametric visualisation of the work's versatility in collaboration with other musicians

The yellow stripe with a blue gradient and the blue dots correspond to study #4, with Aki Onda. My interaction with him did not really happen at a level of expression or phrasing, yet the sonic construction changed at a structural level as well. Aki's sounds unfolded very slowly, creating steady and progressive continuities, varying in density. Accordingly, the blue/ yellow stripe shows greater concentration of sonic continuities. The blue dots show that this collaboration would alter the semantics of my audio-visual performance work. Aki's sonic continuities would reinforce the environmental dimension, weakening the expressive. The vintage quality of his sound materials would reinforce the informational dimension as well.

The yellow stripe with an orange gradient on sonic discontinuities and the yellow dots on semantics correspond to case studies #2 and #3, with Helena Espvall and Tó Trips; the parameterisation of those collaborations is equal to the one of my solo performances, because the dominant type of sonic discontinuities and the semantics of the work are equal to when I play solo. The collaboration with Helena (study #2) led to complex orchestral phrasing. We picked up on each other's chromatic details, creating Gestalts of good continuation at the textural level. With Tó (study #3), the musical outcome was quite different. Similarly to Helena, he is expert in creating Gestalts of good continuation, but the musical motion did not increase in complexity. Tó played with large silences, which made me slow down the musical pace, and unfold longer musical phrasings.

The case studies show that my musical forms can convey a diversity of sonic constructions. Their resilience at the expressive level permits an adaption to different paces, and different types of musical phrasings. It can lead to the discovery of creative potentials that are consonant with the spirit of my musical forms. Their structural flexibility also permits an adaptation to situations in which another musical idiom counteracts that spirit. Exploring creative tensions as creative material can transform the overall semantics of my music, and by extension, the semantics of my audio-visual performance work.

The parametric model of the thesis cannot represent my work's versatility in detail, given its level of abstraction. However, that limitation is useful for representing versatility in a clear and simple way. Whilst the non-versatile parameters assure that the work conveys a characteristic type of experience, the versatile parameters assure the on-going discovery of creative possibilities.

6.5 Summary of Contributions

The research of this thesis gleaned how an audio-visual instrument can convey sonic expression, and how the relations between the sounds can be fully experienced while the image generates additional effects and meanings. The notion of instrument as an open-ended system is crucial in this work, because it embraces instrument, composition and performance at once. My investigations led to three creative principles and a parametric visualisation model, which informed the development of an audio-visual instrument and performance language. The principles concern the sound, the image and the audio-visual relationship. They convey a particular artistic sensibility, but the perceptual approach they are driven from is useful to any audio-visual practice. The model can be used to analyse sonic expression, sensory dominance and spatial presence in any audio-visual performance, revealing how the work converges and diverges from the principles. The previous sections of this chapter discussed the relation between the contributions of this research: three creative principles, a parametric visualisation model, and my practical work. This closing section summarises that relation, so as to facilitate the reader's assessment of the research as a whole.

6.5.1 Sensory Dominance: Summary of Contributions

The parameterisation of sensory dominance is a key contribution to the artistic debate on the subject. It is validated with my extrapolations from neuroscience and psychology, which led to a taxonomy of continuities/ discontinuities related with attention, and to a study on audio-visual mapping and perception. Using these methods, I concluded that vision dominates over audition whenever the visual dynamics exhibit radical discontinuities, and/or the audio-visual relationship produces conclusive concepts of causation. Whilst radical discontinuities attract automatic attention, conclusive concepts lead perception to prioritise fitting information over non-fitting information. I demonstrated how concepts can remain inconclusive, introducing the notion of a fungible audio-visual relationship. In the parametric model, sensory dominance can be inferred from the axes *Visual Discontinuities* and *Ecological Fit*. Those axes can represent the creative principles of visual continuity and audio-visual fungibility, whose exploration allows for perceptual resolution to increase and decrease according to complex sonic constructions.

Every work dispensing with radical visual discontinuities converges with the principle of visual continuity. My work explores thresholds of perceptual simplification, exposing a gray area. The image conveys Gestalts of invariance, exhibiting progressive continuities and ambivalent discontinuities; I never sensed a radical visual discontinuity. Yet, the timescale of my visual apprehensions is much shorter than the duration of a performance, and sustained visual attention could lead discontinuities to become intense. By using the model to analyse performances by other artists, the thesis exposed that the principle can be explored in more straightforward ways, with different purposes; for example, steady and progressive continuities can facilitate the visualisation of symbolic systems, when the intention is to foster semantic perception. My analyses also showed that the model can reveal when sensory dominance is not a creative question because the logics of sonic and visual dynamics are one and the same.

Every audio-visual relationship exhibiting medium ecological fit converges with the principle of fungibility. Since we can form conclusive concepts of causation in spite of inconsistent sensory information, inconsistency does not suffice to characterise medium fit: the audio-visual relationship must create a sense of causation, and confound the base cause and effect relationships. The thesis presented a study that demonstrates how this is achieved by combining synchronised and non-synchronised audio-visual components, with complexity so as to generate confusion. My practical work explores the principle with a fungible mapping and a fungible audio-visual relationship in space; the principle informs instrument design/ configuration and physical set-up. By using the model to analyse performances by other artists, the thesis showed that the principle can be also be explored with consistent audio-visual synchrony. Namely, when understanding cause-effect relationships requires understanding complex symbolic systems, or when the performer's physical gesture is solely occasionally synchronised with the audio-visual mapping. The thesis also showed that the model can indicate why fungible relationships may be undesired in some cases, even when the logics of sound and image are not the same: if the performer is not visible, clearer cause-effect relationships can make the audience understand that the work is being made in real-time.

My instrument and physical setup address the principle of audio-visual fungibility, whether I play solo or in collaboration with other musicians. The question of whether it equally complies with the principle of visual continuity might depend on the perspective of analysis. It does from my creative perspective, considering the timescales of my studio work, and my reasoning about why the visual motion does not counteract the music at larger timescales. Comparing this creative perspective with the audience's experience was beyond the aims of the thesis, inclusively because user-studies or neuro-imaging results would not suffice for a satisfactory conclusion. In any case, the image is projected upon the performer(s), and the digital 3D world appears "alive". It is affected by acoustic input, in multiple ways. There are also visual components whose dynamics do not depend on input detection, several overlapping images with transparencies, and light/ shadow effects. I feel that the visual changes do not obfuscate the experience of music, but they are definitely not secondary. The audio-visual relationship creates what Chion called added value, to describe the expressive and informative value with which a sound affects an image [1994]. By producing a sense of causation, the work emphasises the transformative power of technology. By confounding the base cause and effect relationships, it invites the audience to shift their attention away from technology, and enjoy the experience of multistable percepts and interpretations. The image works as a moving painting. Its role is fundamentally different in other works exploring music with 3D environments, intended for participatory interaction.

6.5.2 Sonic Expression: Summary of Contributions

Beyond the question of sensory dominance, perception science was also useful to formulate a notion of sonic expression. This notion considers how a performer's interaction with their instrument or system informs the sonic dynamics, and the semantics of music. Dynamics and semantics do not distinguish this notion of expression from other notions, but interaction is a distinguishing factor: the interface should convey an awakening of unused abilities. To parameterise interaction I proposed a notion of effort related to cognitive demand and musical timing. By using the parameter *Real-Time Effort in Sound Organisation*, the parametric model also summarises information from other parametric visualisation models, focused on interaction. This parameter can represent the notion of musical expression of the thesis, and by extension, represent my creative principle of sonic complexity, which implies medium to medium-high effort: control and unpredictability should be thresholded so as to convey the performer's idiosyncratic expression.

My creative strategies generate sonic complexity by exploring disparities between human perception and digital analysis as creative material. They led to the discovery of a musical language that explores chromatic shifts, while acoustic and digital sounds at times converge in unison, and other times diverge in independent directions. These creative strategies emphasise a difference between the direct, physical interaction with an acoustic instrument, and the indirect interaction with software, which is mediated through code. The interaction with my instrument entails medium effort. Its behaviours are complex, but a sense of immediacy conveys musical timing, and technical configurations rule out undesired outcomes. By using the model to analyse performances by other artists, the thesis put in evidence that the principle of sonic complexity can be explored in many different ways - including when the interaction requires greater cognitive effort, and even, when timing is not under the performer's direct control.

The idea of digital instruments as epistemic tools (see chapter 2.3), Ryan's notion of local time (see chapter 2.5.2) and the notion of flow in music (see chapter 2.5.4) are useful to situate my particular ways of exploring sonic complexity. By using several layers of digital analysis as creative material, my work explores the epistemic dimension of software so as to create a separation between performer and instrument, rather than a sense of immediacy. As the digital sounds create certain unpredictability, the instrument challenges my body knowledge, as well as the body knowledge of any collaborating musician. Perception science provides a useful perspective over the role of "chance" and effort in the awakening of unused abilities. My immediate interaction with the zither allows for a sense of control over "the time we make, enacted time, dense and polyvalent, the most elaborate aspect of time in music" [Ryan, in Sa et al. 2015]. Whether I'm playing solo or in collaboration with other musicians, the design of the instrument and my compositional strategies allow for the music to reveal that "before it happens, I know when a beat should come", and "I know after, when it didn't" (ibid). That knowledge substantiates in the sonic dynamics, as expression.

6.5.3 Spatial Presence: Summary of Contributions

Regarding the audience's sense of spatial presence, my work intertwines the physical and the digital 3D space through the combination of sonic/visual materials, audio-visual mappings, physical setup and sound spatialisation. The thesis adopted the notion of spatial presence as cognitive feeling, related with semantics and attention. The parametric model uses three discrete points to characterise the *Performative Arena*: *Local, Distributed, Extended*. In this way, it can represent how the mediated space of an audio-visual performance may draw a focus to the performer, to the environment, or to imaginary spaces beyond the physical performance space. As a high level parameter, the *Performative Arena* also provides information that the other parameters in the model do not specify. In this way, spatial presence can be inferred from the model as a whole.

Expanding a characterisation of sounds to the audio-visual domain, the model considers that the *Performative Arena* can be extended through the *Informational* load of the sound and the image (which can evoke other spaces), and the audio-visual relationship (which can explore perceptual cues so as to convey spatial presence in a virtual space). Also, the model considers *Expressive* and *Environmental Semantics* inversely proportional. It draws from literature about the semantics of physical setup in electronic music performance, as well as from my taxonomy of continuities/discontinuities related with attention. By defining musical gesture as musical motion, I stressed the inextricable relation between the semantics and the dynamics of any sonic construction. The relation is equivalent for the image and the audio-visual relationship: whilst discontinuities convey expressive semantics, continuities convey environmental semantics.

6.5.4 Practical Usefulness of the Parametric Model

I used the parametric model to discuss my practical work, and my practical work to discuss the model. Each parameter can summarise several aspects of the work, and one can also use the model to analyse each aspect independently. As an example, I used the model to estimate how a set of musical collaborations would affect my audio-visual performance work as a whole. The visual dynamics, ecological fit and physical setup are similar in every performance, solo or collaborative. Yet the dynamics and semantics of sonic constructions are variable when I play with other musicians. My taxonomy of continuities and discontinuities facilitated the distinction of an expressive and a structural dimension in music, which clarified the model's level of abstraction. By representing the work's versatility at a musical level, the model has put in evidence how the work assures a particular type of experience, and simultaneously enables the discovery of new creative possibilities.

The model cannot represent the details of my practical work, nor can it represent its versatility at an expressive level, given its level of abstraction. Nevertheless, the model can visualise its structural versatility in a simple, useful way. It can be used to analyse my existing instrument, create new audio-visual performances, and develop new audio-visual systems. Beyond personal practice, the model provides an operational means to analyse the relation between sonic expression, sensory dominance and spatial presence in any audio-visual performance language.

Animating the model is beyond the aims of the thesis, but the idea points to future, exciting research possibilities. For example, one could use an animated version to analyse how a same set of instrument behaviours might create different relations of continuity and discontinuity when experienced for different clock-time periods. One could equally use an animated version to predict how the intertwining of continuities and discontinuities drives attention in a large compositional structure. Indeed, an animated version would be useful to analyse how attention might change over time.

CONCLUSION

The questions of this research emerged when I started developing the first version of my audio-visual instrument. The combination of acoustic and digital components raised questions about interaction, and I wanted to extract musical meaning from those questions. I wanted the image to function as a reactive environment, and that was problematic because I had the strong impression that vision tends to subordinate audition. The present research developed a scientific perceptual approach to instrument design, composition and performance, which can be useful to a broad range of researches and creative concerns. It led to three creative principles and a parametric visualisation model that can be used to analyse any audio-visual performance work. The investigations fed back into my artistic work, and will continue to do so into the future.

The literature review contextualised the research and grounded its methods. Looking at certain notions of instrument and spatial presence gave reasons for an approach to the instrument, composition, and performance as a whole. To situate my instrument, I pinned down the difference between physical and digital mediation as well as the technical paradigm with 3D software. That paradigm uses perceptual cues to convey a sense of immediate interaction, but perceptual cues can be explored in different ways, for different reasons. The idea of spatial presence as cognitive feeling related to attention gave reasons for considering how the sound, the image, the audio-visual relationship and the physical setup inform the performative arena, driving the audience's experience.

My concern with visual dominance led me to review different types of audio-visual relationships and the artistic debate on sensory dominance. After distinguishing between "one-to-one" and "structural" approaches to audio-visual performance and animation, I looked at works exploring music with digital 3D environments. This made clear that the one-to-one approach might be adequate for networked participatory interaction, but it is problematic for the experience of music. The question of whether and how visual dominance can be avoided remained unanswered, though. I realised that in audio-visual theory and music the debate lacks resolution.

Perception science established the foundations of my approach to the problem, shedding light upon how perception works, and how multisensory integration relates to attention. It revealed that sudden visual discontinuities and high ecological fit lead to visual dominance, but synchrony in itself does not. This made clear that I needed methods to analyse the sonic and visual dynamics, as well as the audiovisual surplus. I found an inspiring notion of intensity in music perception research, which became useful for the design of those methods.

Perception science also provided a useful perspective over interaction design and sonic expression. After looking at the notion of embodiment, I questioned the relation between intensity, effort and chance. Analysing how different notions of flow inform different types of interaction design showed that effort is a variable that can be used to parameterise different understandings of expression.

I drew from the literature review so as to define the creative principles of sonic complexity, visual continuity and audio-visual fungibility, which are applicable in instrument design, composition and physical performance setup. These principles derive from methods for the analysis of sonic expression and sensory dominance, which later informed the design of the parametric model and the development of my practical work.

Bridging perception science and philosophy enabled me to clarify how intention frames the mind through previous assumptions, and intentionality conveys the experience of complexity. I examined how interfaces convey one or the other, elaborating a notion of musical expression where the combination of effort and embodiment conveys intentionality and sonic complexity. This understanding grounded the principle of sonic complexity. Considering different notions of expression, I established a method for the analysis of any interaction design.

The question of sensory dominance required a method to analyse the sonic and visual dynamics, and compare their relative strength. The method should enable the distinction between apprehensions automatically driven through stimuli, and apprehensions more dependent on individual control. This led me to redefine “intensity” and create a taxonomy of continuities and discontinuities related to attention, which was illustrated with examples from audio-visual performance and music. The taxonomy grounded the principle of visual continuity.

The next challenge was to understand how an audio-visual relationship can produce a sense of causation without subordinating audition. Bridging perception science and audio-visual theory, I created a method to analyse how different ratios of congruence/ incongruence affect perceptual prioritisation. I introduced the notion of a fungibile audio-visual relationship, which produces inconclusive concepts of causation. The subsequent study demonstrated that a fungible mapping combines synchronised and non-synchronised components, exhibiting complexity and inconsistency. The study demonstrated my distinction between low, medium and high ecological fit, which is equally applicable to the audio-visual relationship in space.

The research proceeded with the design of the parametric visualisation model. I begun by looking at other models, so as to find inspiration and learn from problems. Using a limited amount of parameters, my model should be applicable to any technical platform, aesthetical approach and physical setup. To reveal how a work might converge and diverge from my creative principles, the model required axes for real-time effort, sonic & visual discontinuities and ecological fit. To indicate spatial presence, the model also required axes for physical setup and semantics. Those axes would not suffice to represent one's sense of presence beyond the physical space, and that was challenging because the model should not attempt to parameterise experience in itself. I decided to include a high-level parameter for the performative arena, which enables the disambiguation of information that the other axes do not represent directly. Using the model to analyse different audio-visual performance works showed that my three creative principles can be explored in many different ways, and that the model is useful whether visual dominance is a creative concern or not.

Describing the first version of my instrument and physical setup through the lens of the creative principles and the parametric model was very useful to clarify artistic motivations. Now that the boundaries of my previous my insights were clear, I was prepared to specify new software from scratch. The principles and the model framed and inspired the expansion of my creative strategies in depth. The following audio software used several layers of digital analysis to increase sonic complexity. I adopted a particular zither tuning system, and discovered the potentials of applying the difference between the detected pitch and the closest tone/half tone to pre-recorded sounds, as a pitch-down value. I developed specific combinations of zither techniques and digital sounds, creating digital constraints that enable levels of unpredictability and rule out undesired outcomes. My design strategies enabled versatile musical dramaturgies, intertwining different semantic dimensions. The music became much richer and meaningful than with the first version of the instrument.

The audio software was then combined with a 3D engine, written from scratch. The creative principle of visual continuity rules out radical visual discontinuities, and that was an inspiration to overlap multiple images with transparencies, in ways that accentuated the impression of an organic, moving painting. The Gestaltist psychology was useful to create a wealth of visual discontinuities that facilitate perceptual simplification, but my creative decisions were not driven by demonstrative purposes. Exploring gray areas where the intensity of visual dynamics might depend on time and individual predispositions was more exciting than assuring visual continuity. And from a scientific perspective, placing perceptual simplification in question stressed how the threshold between ambivalent and radical discontinuities depends on the perspective of analysis. Furthermore, the principle of fungibility inspired a complex audio-visual mapping, which uses the difference between the detected pitch and the closest tone/half tone as creative material; the mathematical value is applied as delay time, and it also influences the magnitude of visual changes. The digital 3D world became much more organic than in the first version of my instrument. The semantics of the image continued highly environmental, but attention was invited to focus on a greater amount and variety of ambivalent discontinuities. Specifying the software from scratch also gave me a greater control over how the visual dynamics affect the sound; I explored perceptual cues for direction and distance in ways that convey the music. I also worked on the physical configuration of the instrument, improving sound quality so as to stress the distinguishing characteristics of the zither and the digital sounds. I continued to use digital 3D sound with a double, inverted stereo system, considering the principle of audio-visual fungibility; the work improved with the addition of a particular compressor, reverb and amplifier.

To conclude the research, I discussed its contributions considering its questions and aims, as well as the relation between my practical work and the parametric visualisation model. I discussed how the model parameterises sonic expression, sensory dominance and spatial presence, considering my creative principles and how this research can contribute to the scientific creative community. Then I used the model to represent my solo audio-visual performance work, detailing how each parameter summarises a set of different factors. This raised the question of whether, and how, the model could represent the versatility of my work in collaboration with other musicians. It also raised the question of whether, and how, the model could be used to estimate how those musical collaborations would affect the audio-visual performance work as a whole. I found a way to clarify the model's level of abstraction, by using my taxonomy of continuities and discontinuities to distinguish a structural and an expressive dimension in music. The distinction became useful in four case studies, where I used the taxonomy to analyse very different musical outcomes, and the model to estimate how they would affect my audio-visual performance work. Indeed, the model can represent how some aspects of my work are versatile, and others are not.

The research gleaned how an audio-visual instrument can convey sonic expression, and how the image can create a surplus of meaning without obfuscating the relations between the sounds. The research was motivated by personal artistic concerns, yet its perceptual approach can be useful to any audio-visual practice. Indeed, the relation between sonic expression, sensory dominance and spatial presence is crucial in any audio-visual performance, regardless of its particular sphere of creative concerns and aesthetical approach.

Bibliography

___ *Immersed in Media: Telepresence Theory, Measurement & Technology*; ed. M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, R. Schaevitz, Springer International Publishing (2015):73-94

___*The Handbook of Multisensory Processes*. Ed. G. A. Calvert, C. Spence and B. E. Stein, MIT Press, 2004.

___*Sons & Lumieres – Une histoire du son dans L'art du 20e siècle*, Centre Pompidou, Parcours Expositions, 2005.

Barri, Tarik, official site about the *Versum* project, http://tarikbarri.nl/projects/versum (Accessed August 30, 2016)

Bergson, H.L., *Matter and Memory*, transl. N.M. Paul, W.S. Palmer, Allen, London, 1910.

Bertelsen, Olav W., Morten Breinbjerg, and Søren Pold, "Emerging Materiality: Reflections on Creative Use of Software in Electronic Music Composition." *Leonardo* 42*, no. 3* (2009): 97–202

Bertelson, P.; Vroomen, J.; de Gelder, B.; Driver, J., "The Ventriloquist Effect Does Not Depend on the Direction of Deliberate Visual Attention". In *Perception and Psychophysics 62, no. 2* (February 2000): 321–332.

Berthaut, Florent; Katayose, Haruhiro; Wakama, Hironori; Totani, Naoyuki; Sato, Yuichi, "First Person Shooters as Collaborative Multiprocess Instruments". In *Proceedings of the International Conference on New Interfaces for Musical Expression* (2011): 44–47

Birnbaum, D.; Fiebrink, R.; Malloch, J. and Wanderley, M. M., "Towards a Dimension Space for Musical Devices". In *Proceedings of the International Conference on New Interfaces for Musical Expression,* Vancouver, BC, Canada (2005)

Boashash, Boualem, *Time Frequency Analysis*, Gulf Professional Publishing, 2003.

Boltz, Merilyn, "The Cognitive Processing of Film and Musical Soundtracks". In *Memory & Cognition 32* (2004): 1194-1205.

Brakhage, Merilyn "On Stan Brakhage and Visual Music," *Vantage Point* (January 31, 2008), http://vantagepointmagazine.wordpress.com/2008/01/31/on-stan-brakhage-and-visual-music (accessed September 23, 2012).

Bregman, A. S., *Auditory Scene Analysis – The Perceptual Organization of Sound,* Cambridge: MIT Press, 1990

Broadhurst, Susan, *Digital Practices: Aesthetic and Neuroesthetic Approaches to Performance and Technology*, New York, Hampshire, Palgrave Macmillan, 2007

Brougher, K., Strick, J., Wiseman, A. and Zilczer, J., *Visual Music: Synaesthesia in Art and Music Since 1900*. Thames and Hudson, 2005

Brown, S.W. and Boltz, M., "Attentional Processes in Time Perception: Effects of Mental Workload and Event Structure". In *Journal of Experimental Psychology: Human Perception and Performance*, 28 (2002): 600-615.

Cage, John, *Silence: Lectures and Writings 1939 – 1961*, Wesleyan University Press, Connecticut, 1961.

Cannon, Joanne, and Favilla, Stuart, "The Investment of Play: Expression and Affordances in Digital Musical Instrument Design". In *ICMC proceedings*, Ljubljana (2012), 459-466.

Ciciliani, Marko, "Towards an Aesthetic of Electronic-Music Performance Practice". In: *Proceedings of the ICMCISMC*, Athens, Greece (2014)

Ciciliani, Marko and Mojzysz, Zenon, "Evaluating a Method for the Analysis of Performance Practices in Electronic Music". In *Proceedings of the Second International Conference for Live Interfaces* (2015), ISBN 978-989-746-060-9

Chen, Jenova, "Flow in Games" (2006), http://www.jenovachen.com/flowingames/foundation.htm, (accessed January 20, 2013).

Chion, Michel, *Audio-Vision: Sound on Screen*. Trans. C. Gorbman, New York, Columbia University Press, 1994.

Cascone, Kim, "The Esthetics of Failure: Post-Digital Tendencies in Computer Music". In *Computer Music Journal* 24 no. 4 (2000): 12-18.

Colavita, F. B., "Human Sensory Dominance". In *Perception & Psychophysics 16, no. 2* (1974), 409-412.

Collins, N., McLean, A., Rohrhuber, J. & Ward, A., "Live Coding Techniques for Laptop Performance". In *Organised Sound* 8 no. 3 (2003): 321-30.

Conrad, Tony, blog about *The Flicker*, http://flicker75.blogspot.pt/2008/01/flicker.html (Accessed August 30, 2016)

Conrad, Tony, excerpts of *The Flicker*, https://www.youtube.com/watch?v=ZJbqnztjkbs (Accessed August 30, 2016)

Cowan, Nelson, "The Magical Number 4 in Short-Term Memory: a reconsideration of mental storage capacity". In *Behavioral and Brain Sciences 24*, no. 1 (2001): 129- 130.

Csikszentmihalyi, Mihaly, *Creativity: Flow and the Psychology of Discovery and Invention*. HarperCollins Publishers, 1996.

Csikszentmihalyi, Mihaly, and Csikszentmihalyi, Isabella S., *Optimal Experience: Psychological Studies of Flow in Consciousness,* Cambridge University Press, 1988.

Cytowic, R.E., *Synesthesia: A Union of the Senses (2nd ed),* Cambridge, MIT Press (1st ed 1989, Springer-Verlag), 2002.

Draper, J. V., Kaber, D. B., & Usher, J. M., "Telepresence". In *Human Factors 40*, no. 3 (1998): 354–375.

Dourish, Paul, *Where the Action Is: The Foundations of Embodied Interaction*, MIT Press, 2004.

Dyson, Frances, *Sounding New Media – Immersion and Embodiment in the Arts and Culture*, University of California Press, 2009.

Emmerson, Simon, *Living Electronic Music*, Hampshire, U.K.: Ashgate Publishing Limited, 2007.

Fishinger, Elfriede, performance with the Oscar Fishinger's *Lumigraph,* https://vimeo.com/13787280 (Accessed August 30, 2016)

Fishinger, Oscar, *Motion Painting no.1*, https://new.vk.com/video2804234_168196055 (Accessed August 30, 2016)

FMOD audio library, "blog", http://www.fmod.org/forum/, official website (accessed April 2013)

Freed, A., Chaudhary, A. and Davila, B. "Operating Systems Latency Measurement and Analysis for Sound Synthesis and Processing Applications". In ICMC Proceedings (1997): 479–81.

Gaver, W. W., "Auditory Icons: Using Sound in Computer Interfaces". In *Human-computer Interaction 2* (1986): 167–177.

Gilakjani, Abbas P., "Analysis of Working Memory and Its Capacity Limit in Visual and Auditory Information". In *Theory and Practice in Language Studies 2*, no. 2 (2012): 393-397.

Grierson, Mick, "Noisescape: A 3d Audiovisual Multi-User Composition Environment". In *The World as Virtual Environment*, ed. J. Birringer, T. Dumke, K. Nicolai, Dresden: Trans-Media-Akademie Hellerau (2007): 160-168.

Grimes, Tom, "Audio-Visual Correspondence and its role in Attention and Memory". In *Educational Technology in Research and Development 38, no 3* (1990): 15-25.

Gritten, Anthony, "Drift". In *Music and Gesture*, ed. Anthony Gritten and Elaine King, Ashgate Publishing Limited (2006) :104-125.

Hamilton, R. "q3osc: Or How I Learned to Stop Worrying and Love the Game". In *Proceedings of the International Computer Music Association Conference 2008* (a).

Hamilton, R., "Maps and Legends: Designing FPS-Based Interfaces for Multi-User Composition, Improvisation and Immersive Performance". In *Computer Music Modeling and Retrieval. Sense of Sounds: 4th International Symposium 2007*, Revised Papers, 2008 (b).

Hartmann, T., Wirth, W.,Vorderer, P., Klimmt, C., Schramm, H., ·Böcking, S., "Spatial Presence Theory: State of the Art and Challenges Ahead". In *Immersed in Media: Telepresence Theory, Measurement & Technology*; ed. M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, R. Schaevitz, Springer International Publishing (2015):115-135.

Hatten, Robert, "A Theory of Musical Gesture and its Application to Bethoven and Schubert". In *Music and Gesture*, ed. Anthony Gritten and Elaine King, Ashgate Publishing Limited (2006) :142-164.

Hill, Gary, *Black and White Text*, 1980 https://www.youtube.com/watch?v=bg1O3NcPwBg (accessed July 2015).

Hunicke, R. and Chapman, V., "AI for Dynamic Difficult Adjustment in Games". In *Proceedings of the Challenges in Game AI Workshop, 19th National Conference on Artificial Intelligence* (2004).

Ikeda, Ryoji, *Superimposition*, 2012, https://vimeo.com/49873167 (accessed July 2015).

Ikeda, Ryoji, *Test Pattern*, 2008, https://www.youtube.com/watch?v=GPhAvyrZ28o & http://www.ryojiikeda.com/project/testpattern/ (accessed July 2015).

Jentzsch, Wilfried and Ishii, Hiromi, "Multimedia "Visual Music": Analogical Processes Between Images and Sounds". In *Proceedings of the third EMSAN-CEMC Colloquium, "Round Table Discussion and Forum of Development about Asian Electroacoustic Music"*, 2010. www.omf.paris-sorbonne.fr/EMSAN/2010/ISHII-JENTZSCH-2010.pdf

Johnston, Andrew, "Beyond Evaluation: Linking Practice and Theory in New Musical Interface Design". In *Proceedings of the International Conference on New Interfaces for Musical Expression* (2011), 280–283.

Karkowski, Zbigniew, "The Method Is Science, the Aim Is Religion" (1992). http://www.desk.nl/~northam/oro/zk2.htm (accessed December 17th, 2013)

Kelkar, Tejaswinee and Indurkhya, Bipin, "A Motion-Based Approach to Analysis, Notation and Creation of Hindustani Music". In *XcoAx Proceedings* (2014), 222–233.

Kennerly, David Ethan, "Fun is Fine: Toward a Philosophy of Game Design" (June 22, 2003) http://www.finegamedesign.com/fun_is_fine.html (accessed September 5th 2012)

Kiefer, Chris, "Interacting with Text and Music: Exploring Tangible Augmentations to the Live Coding Interface". In *Proceedings of the Second International Conference for Live Interfaces* (2015), ISBN 978-989-746-060-9

Kilbourn, K, and Isaksson, J. "Meaning through doing: The role of affordances over time" In *Sixth Nordcode Seminar & Workshop, Design Semiotics in Use* (2007): 6-8.

Knudsen, E. I., "Fundamental Components of Attention". In *Annual Review of Neuroscience 30, no.1* (2007), 57-78.

Kobayashi, M.; Osada, Y.; Kashino, M. "The Effect of a Flashing Visual Stimulus on the Auditory Continuity Illusion". In *Perception & Psychophysics* vol. 69 no. 3 (2007): 393-399.

Koffka, K., *Principles of Gestalt Psychology,* London: Lund Humphries, 1935.

Koster, Ralph, *Theory of Fun*, Paraglyph Press, 2004. Info and excerpts including foreword by Will Wright at http://www.theoryoffun.com, (accessed January, 2013)

Kubovy, M. and Schutz, M., "Audio-Visual Objects". In *Review of Philosophy and Psychology 1, 1,* Springer Science and Business Media B.V. (2010), 41–61.

Kubovy, M. and Yu, M., "Multistability, Cross-Modal Binding and the Additivity of Conjoined Grouping Principles". In *Philosophical Transactions of the Royal Society Vol. B, no. 367* (2012) :954–964. doi: 10.1098/rstb.2011.0365

LeGrice, Malcolm, *Abstract Film and Beyond*, MIT Press, 1982.

Lerdahl, F.; Jackendoff, R. *A Generative Theory of Tonal Music*, Cambridge MA, MIT Press, 1983.

López, Francisco, official Web Site, “Against the Stage” (February 2004), http://www.franciscolopez.net/stage.html (accessed September 19, 2012).

McDonald, J.J, Störmer, V.S., Martinez, A., Feng, W. and Hillyard, S.A., “Salient Sounds Activate Human Visual Cortex Automatically”. *In Journal of Neuroscience 33 no. 21* (2013): 9194-9201. doi: 10.1523/JNEUROSCI.5902-12.2013

McLaren, Norman, *Dots* (1940), http://vimeo.com/15919138 (accessed April 2013)

Machover, Tod, “Beyond Guitar Hero - Towards a New Musical Ecology.” In *RSA Journal*, London (2009).

Magnusson, Thor, and Mendieta, Enrike H., “The Acoustic, the Digital and the Body: a Survey on Musical Instruments.” In *NIME ’07: Proceedings of the 7th International Conference on New Interfaces for Musical Expression*, New York, NY, USA, ACM (2007) :94–99

Magnusson, Thor “Of Epistemic Tools: Musical Instruments as Cognitive Extensions”. In *Organised Sound 14, no. 2* (2009): 168–176.

Magnusson, Thor, “An Epistemic Dimension Space for Musical Devices”. In *Proceedings of the Conference on New Interfaces for Musical Expression,* Sydney, Australia (2010)

Magnusson, Thor, “Designing Constraints: Composing and Performing with Digital Musical Systems”. In *Computer Music Journal 34*, no.4 (2010): 62–73.

Magnusson, Thor, “Scoring With Code: Composing with Algorithmic Notation”. In *Organised Sound 19*, no.4. (2014) 268-275. ISSN 1355-7718

Magnusson, Thor, *Fermata,* Video footage from performance at the ICLI 2014 conference, https://drive.google.com/file/d/0B9f9deWrFgQuSU1NbFJmdkZWMU0/view?usp=sharing and rendering of the *Threnoscope* system https://vimeo.com/75380587 (accessed April 2015).

Mäki-Patola, T and P. Hämälainen, P., “Latency Tolerance for Gesture Controlled Continuous Sound Instrument Without Tactile Feedback”. In *ICMC Proceedings* (2004).

Mandelbrot, B., *The Fractal Geometry of Nature*. W.H. Freeman and Company, 1982.

Marmelade game engine, “Questions”, official website, https://devnet.madewithmarmalade.com/index.html (accessed April 2013).

Marmelade game engine, “Community”, official website, http://www.madewithmarmalade.com/devnet/forums/marmalade-community-code-projects (accessed April 2013).

Maximillian audio library, http://maximilian.strangeloop.co.uk/, official website (accessed April 2013)

McGurk, H.; McDonald, J. "Hearing Lips and Seeing Voices". In *Nature 264* (December 23, 1976), 746–748, http://www.nature.com/nature/journal/v264/n5588/abs/264746a0.html (accessed September 23, 2012)

Medienkunstnez, site about Tony Conrad's *Flicker*, http://www.medienkunstnetz.de/themes/overview_of_media_art/perception/5/ (Accessed August 30, 2016)

Moritz, William, *The Dream of Color Music, And Machines That Made it Possible*. Animation World Magazine, 1 April 1997. http://www.awn.com/mag/issue2.1/articles/moritz2.1.html (Accessed August 30, 2016)

Mudd, T., Holland, S., Mulholland, P. and Dalton, N., "Nonlinear Dynamical Systems as Enablers of Exploratory Engagement with Musical Instruments". In *Proceedings of the Second International Conference for Live Interfaces* (2015), ISBN 978-989-746-060-9

Murray-Browne, Tim, Di Mainstone, Bryan-Kinns, Nick, D. Plumbley, Mark. "The Medium is the Message: Composing Instruments and Performing Mappings". In *Proceedings of the International Conference on New Interfaces for Musical Expression* (2011), 56–59.

Niblock, Phill, *Movements of People Working* DVD, Microcinema, 2003.

Niblock, Phill, *The Magic Sun* DVD, with Sun Ra, Marshall Allen, John Gilmore and Danny Davis, 1966. http://www.dailymotion.com/video/xmmkj0_sun-ra-magic-sun_shortfilms (accessed August 10, 2015).

Noë, Alva, *Action in Perception*. MIT Press, Cambridge, 2004.

Onda, Aki, official website, http://www.akionda.net/ (accessed August 10, 2015).

Pachet, F., "Enhancing Individual Creativity with Interactive Musical Reflective Systems". In *Musical Creativity: Current Research in Theory and Practice*, ed. I. Deliège and G. Wiggins, Psychology Press (2004).

Palmer, S. *Vision Science, Photons to Phenomenology,* Cambridge, MA, MIT Press, 1999.

Pan Sonic, *Corona* CD, Blast First Petite, 2010 https://www.youtube.com/watch?v=kPhkH1da08o (accessed July 2015)

Pardue, L., Dongjuan, N., Harte, C., and McPherson, A., "Low-Latency Audio Pitch Tracking: a Multi-Modal Sensor-Assisted Approach". In *NIME proceedings* (2014). http://www.nime2014.org/technical-programme/proceedings/ (accessed July 2014)

Paul, Christiane, *Digital Art*, Thames & Hudson, 2003.

Pick, H. L.; Warren, D. H.; Hay, J. C. "Sensory Conflict in Judgements of Spatial Direction". In *Perception Psychophysiology, no.4* (1969), 203–205.

Poepel, C. and Overholt, D. "Recent Developments in Violin-Related Digital Musical Instruments: Where Are We and Where Are We Going?" In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Paris, France, (2006), 390–395

Popper, Karl R., *The Open Universe: An Argument for Indeterminism*, Totowa, Rowland and Littlefield, 1982

Pressing, Jeff, “Some Perspectives on Performed Sound and Music in Virtual Environments”. In *Presence 6, no. 4* (1997).

Pressing, Jeff, “Cognitive Complexity and the Structure of Musical Patterns”. In *Generative Processes in music - The Psychology of Performance, Improvisation, and Composition*, ed. by J. Sloboda, New York, Oxford University Press (1987), 129-178.

Radigue, Elaine, *Triologie de la Mort* CD, Experimental Intermedia, 1998.

Rimington, A. Wallace. *Method of and apparatus for producing color effects*, 1895.

Rimington, A. Wallace. *Colour-music : the art of mobile colour*. Hutchinson, 1912.

Riva, G., Mantovani, F., Waterworth, E., Waterworth, J., “Intention, Action, Self and Other: An Evolutionary Model of Presence”, In *Immersed in Media: Telepresence Theory, Measurement & Technology*; ed. M. Lombard, F. Biocca, J. Freeman, W. IJsselsteijn, R. Schaevitz, Springer International Publishing (2015):73-94

Robinson, C.,· Barnhart, W., Rivera, S., “Auditory Stimuli Slow Down Responses and First Fixations: Support for Auditory Dominance in Adults”. In *37th Annual Meeting of the Cognitive Science Society*, Pasadena, California (2015). DOI: 10.13140/RG.2.1.1564.5286

Roessler, Otto, *Das Flammenschwert oder wie hermetisch ist die Schnittstelle des Mikrokonstruktivismus?* (in German, Benteli, 1996, ISBN 3-7165-1017-3

Rubin, E. *Visuell Wahrgenommene Figuren - Studien in Psychologischer Analyse*. Copenhagen, Gyldendalske Boghandel, 1921.

Ryan, Joel, “Some Remarks on Musical Instrument Design at STEIM”, Contemporary music review, Vol. 6 issue 1 (1991), 3-17. Online at http://www.steim.org/steim/texts.php?id=3 (accessed September 18, 2013)

Sa, Adriana; Ryan, Joel; McPherson, Andrew; Tanaka, Atau; Magnusson, Thor; Van der Heide, Edwin; Grierson, Mick; Carvalhais, Miguel; McLean, Alex, “Live Interfaces: Problematising Foundations”. In *Proceedings of the Second International Conference for Live Interfaces* (2015), ISBN 978-989-746-060-9

Sacau, A., Laarni, J., Hartmann, T., “Influence of Individual Factors on Presence”. In *Computers in Human Behavior 24, 5*, Elsevier Science Publishers B. V. (2008), 2255-2273

Schaeffer, Pierre, *Traité des Objets Musicaux: Essai Interdisciplines*. Éditions du Seuil, Paris, 1966.

Schorno, Daniel and Karlsson, Haraldur *7-of-12 dialectologies* video, https://vimeo.com/27694220 (accessed July 20, 2015)

Schroeder, F. and Rebelo, P., “The Pontydian Performance – The Performative Layer”. In *Organised Sound* 14, 2 (2009)

Schubert, T. “A New Conception of Spatial Presence: Once again, with Feeling”. In *Communication Theory 19*, no. 2 (2009):161–187

Schutz, M., and Kubovy, M., “Causality and Cross-Modal Integration”. In *Journal of Experimental Psychology: Human Perception and Performance 35, 6* (2009), 1791–1810

Shams, L., Kamitani, Y., & Shimojo, S., “Visual illusion induced by sound”. In *Cognitive Brain Research 14* (2002), 147-152

Sinnett, S.; Spence, C.; Soto-Faraco, S. “Visual dominance and attention: The Colavita Effect Revisited”. In *Perception & Psychophysics 69, no. 5* (2007), 673-686.

Smalley, D., “*Spectro-morphology and Structuring Processes”*, in Emmerson, S. (ed.) *The Language of Electroacoustic Music*, London, Macmillan (1986), 61-93.

Snyder, Bob, *Music and Memory: An Introduction.* Cambridge, MIT Press, 2001.

Sensorband’s official website/ archive, http://www.sensorband.com (accessed September 23rd 2012)

Shanken, Edward A., *Art and Electronic Media*. Phaidon, 2009.

Sorensen, Vibeke, "Global Visual Music Jam Project", *Leonardo 38, No. 4* (2006), 316-316.

Sorensen, Vibeke, Puckette, Miller and Steiger, Rand "Global Visual Music Project" (1997) http://www.visualmusic.org/ (accessed April, 2013)

Spence, Charles, “Explaining the Colavita Effect”. In *Progress in Brain Research 176,* ed. N. Srinivasan, Elsevier (2009): 245-58. doi: 10.1016/S0079-6123(09)17615-X.

Stevens, Meghan, *Music and Image in Concert*. Music and Media, 2009.

Sweetser, P. and Wyeth, P., “GameFlow: a model for evaluating player enjoyment in games”. In *Computers in Entertainment (CIE) - Theoretical and Practical Computer Applications in Entertainment 3, 3* (2005): 3.

Talsma, D.; Senkowski, D.; Soto-Faraco, S.; Woldorff, M. G. “The Multifaceted Interplay Between Attention and Multisensory Integration”. In *Trends in Cognitive Sciences 14, no.9* (2010), 400-410.

Tanaka, Atau, “Sensor-Based Musical Instruments and Interactive Music”, in *Creative and Performance Modes*, ed. Dean, R., The Oxford Handbook of Computer Music , (2009), 233-257

Tenney, James, *Meta Hodos and Meta Meta Hodos: a Phenomenology of 20th Century Musical Materials and a Study of Form.* Oakland, Frog Peak Music, 2nd edition, 1988.

Toiviainen, Petri, “The Psychology of Electronic Music”, in *The Cambridge Companion to Electronic Music*, ed. Nick Collins and Julio Escriván, Cambridge University Press (2008)

Torque game engine, “forum”, official website, http://www.garagegames.com/community/forums (accessed April 2013).

Variable Geometry Orquestra, concert at ZDB, Lisbon 2014, https://www.youtube.com/watch?v=yZT42pMph2g (accessed July 2015).

Vasulka, Woody and Vasulka, Steina, *Heraldic View* video, 1974, http://www.fondation-langlois.org/html/e/page.php?NumPage=481 (accessed August 10th 2015).

Vasulka, Steina, and Vasulka, Woody, *Machine Media.* New York, MoMA, 1996.

Vasulka, Steina, *Strange Music for Nam June Paik* video*,* 2013, https://www.youtube.com/watch?v=9mg1weOHWSs (accessed August 10th 2015).

Vasulka, Steina, "MIDI violin demo" video, 1998, http://www.fondation-langlois.org/html/e/page.php?NumPage=419 (accessed August 10th 2015).

Wassenhove, V. V.; Buonomano, D. V. ; Shimojo, S.; Shams, L., "Distortions of Subjective Time Perception Within and Across Senses". In *PLoS ONE* vol. 3 no. 1 (2008) http://www.plosone.org/article/info:doi/10.1371/journal.pone.0001437 (accessed September 23, 2012).

Waisvisz, M., "Gestural Round Table". STEIM Writings (1999) http://steim.org/media/papers/Gestural%20round%20table%20-%20Michel%20Waisvisz.pdf (accessed July 2014).

Welch, R. B.; Warren, D. H., "Immediate Perceptual Response to Intersensory Discrepancy". In *Psychology Bulletin 88, no. 3* (1980): 638–667.

Vergo, Peter, *The Music of Painting: Music, Modernism and the Visual Arts from the Romantics to John Cage*. Phaidon, 2010.

Wertheimer, Max, "Laws of Organization in Perceptual Forms" (first published as "Untersuchungen zur Lehre von der Gestalt II", in *Psycologische Forschung* 4, 1923: 301-350). Translation published in W.Ellis (1938), *A source book of Gestalt psychology,* 71-88, London: Routledge & Kegan Paul, 1938.

Westcott, M.R., *Toward a Contemporary Psychology of Intuition*. Holt, Rinehart, 1968.

Whitney, John, *Digital Harmony*, McGraw-Hill Inc., US, 1980.

Whitney, John, *Arabesque* video, 1975, https://www.youtube.com/watch?v=w7h0ppnUQhE (accessed August 10th 2015).

Wilfred, Thomas, silent excerpt of the *Clavilux*, https://www.youtube.com/watch?v=ojVX8FWYc4g (Accessed August 30, 2016).

Voegelin, Salome, *Listening to Noise and Silence: Towards A Philosophy Of Sound Art.* Continuum, 2010.

Vorderer, P., Hartmann, T., Klimmt, C., "Explaining the enjoyment of playing video games: the role of competition". In Proceedings of the Second International Conference on Entertainment Computing (2003), 1-9

Wozniewski, Mike, Settel, Zack and Cooperstock, Jeremy, "A Framework for Immersive Spatial Audio Performance". In *Proceedings of New Interfaces for Musical Expression*, 2006 (a).

Wozniewski, Mike, Settel, Zack and Cooperstock, Jeremy, "A Paradigm for Physical Interaction with Sound in 3-D Audio Space". In *Proceedings of International Computer Music Conference* (ICMC), 2006 (b).

Wozniewski, Mike, Settel, Zack and Cooperstock, Jeremy, "A Spatial Interface for Audio and Music production". In *Proceedings of the International Conference on Digital Audio Effects* (DAFx), 2006 (c).

Wozniewski, Mike, Settel, Zack and Cooperstock, Jeremy, *Audioscape* research project, http://wikibin.org/articles/audioscape.html (accessed August 18, 2015)

Young, La Monte and Zazeela, Marian, *Dream House* video (begun in 1962) https://www.youtube.com/watch?v=3ahgq-zVQLc (accessed August 18, 2015)

Zettl, H. *Sight Sound Motion: Applied Media Aesthetics*. Wadsworth Publishing, 3rd edition, 1998.