COMPOSING WITH ISOMORPHIC AUDIOVISUAL GESTALTS

TADEJ DROLJC

A thesis submitted to the University of Huddersfield in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Huddersfield

November 2018

Abstract

This research represents an attempt to explore aesthetic deviations from established audiovisual practices through the composition of audiovisual works based on the use of audiovisual gestalts. Within the submitted portfolio, the visual materials are made from digitally generated 3D graphics rendered with lines and points suspended in negative space whilst the musical side is anchored in post-acousmatic (Adkins et al., 2016) and post-digital (Cascone, 2002) beats-infused aesthetics. The aesthetic deviations were explored through different qualities and behaviors of gestalts that span between the organic and the artificial. Organic qualities were achieved by exploiting physics or pseudo physics-based algorithms within spatio-temporal continuum while artificial qualities were derived from the use of metric grids, step transformations and rectangular shapes. This exploration includes different approaches to the creation of audiovisual material such as 1) exploitation of the motion of 3D graphics for the purposes of sonification, visual effects and temporal distortion, 2) juxtapositions of different audiovisual hierarchies in terms of which medium leads and which follows.

This written text accompanies the portfolio of 1) computer generated audiovisual works composed between 2014 and 2018, consisting of three solo works in total length 36'06" and two collaborative works in total length 7'47", 2) a musical study piece that is 9'29" in length, 3) *Stringvencer* – a sequencer that utilises physics-based algorithms for temporal deformations of pre-composed musical structures, consisting of one *Max* patch and two *Max for Live Devices* 4) two *Max* externals and accompanying patches for optimising workflow when using 3D models within *Max* and for connecting points with lines in a 3D space based on the distances between points, 5) media clips and *Max* patches that support the commentary.

The written text presents theoretical models and concepts as well as tools and techniques that lie behind the construction of the works in the portfolio. Following the presentation of the portfolio in Chapter 2, the commentary continues with the presentation of a theoretical framework in Chapter 3. This theoretical framework, which I named the *Isomorphic Audiovisual Paradigm*, encompasses models of perception that deal with the perceptual integration and segregation of elements within an audiovisual scene. The models are borrowed from gestalt psychology (isomorphism and grouping principles) and psychophysics (binding factors). In Chapter 4 the aesthetic influences and their integration into my audiovisual language are discussed. Chapter 5 focuses on my compositional practice and thinking. The topics of *time*, compositional indeterminacy and what I call *audiovisual hierarchies*, *audiovisual tightness* and *audiovisual escapology* are discussed.

The workflow and the development of a custom-made audiovisual system that represents my working environment are presented in Chapter 6. Chapter 7 discusses the technical solutions to the problems faced in the course of the research.

Table of Contents

Instructions

All submitted files are located on the USB stick. Video files of audiovisual works need to be copied from the USB stick to the computer as they most likely will not run smoothly from the USB stick.

All the excerpts from my works that are referenced in the commentary are hyperlinked in the text. In order to access *Capillaries Capillaries*, *Singing Sand* and *Spaceless Latitudes* online, passwords are needed (see “List of Submitted Pieces” on the next page). All referenced excerpts are from the pieces submitted.

When watching full compositions, *Singing Sand* should be played back from the file provided on the USB stick. The online version is slightly different (darker and few micro edits are missing). However, excerpts of *Singing Sand* can be watched either online or from the file provided.

List of Submitted Pieces

Capillaries Capillaries

Path: Audiovisual Works \ Capillaries Capillaries.mov

www: https://vimeo.com/207005521 PASSWORD: capilalipac

Length: 22′ 42″

Format: Audiovisual composition, UHD 4K, 60FPS, stereo audio

Credits: Tadej Droljc (audiovisual composition), Gregor Zemljič (mastering)

Singing Sand

Path: Audiovisual Works \ *Singing Sand*.mov

www: https://vimeo.com/252644715 PASSWORD: SinSan

Length: 12′ 34″

Format: Audiovisual composition, UHD 4K, 60FPS, stereo audio

Credits: Tadej Droljc (audiovisual composition), Gregor Zemljič (mastering)

Synspecies – Teaser

Path: Audiovisual Works \ Synspecies - Teaser.mov

www: https://vimeo.com/251243760

Length: 1′ 19″

Format: Audiovisual composition, UHD 4K, 60FPS, stereo audio

Credits: Elías Merino (music), Tadej Droljc (visuals), Gregor Zemljič (mastering)

Synspecies - Spaceless Latitudes

Path: Audiovisual Works \ Synspecies – Spaceless Latitudes.mov

www: https://vimeo.com/268028565 PASSWORD: synspaclati

Length: 6′ 28″

Format: Audiovisual composition / Installation, 3 x HD, 60FPS, 4.1 audio (installation version); UHD 4K, 60FPS, stereo audio (portfolio version)

Credits: Elías Merino (music), Tadej Droljc (visuals), Gregor Zemljič (mastering)

LPM Promo Video

Path: Audiovisual Works \ LPM Promo Video.mov
www: https://vimeo.com/256164876
Length: 0’ 58”
Format: Audiovisual composition, UHD 4K, 30FPS, stereo audio
Credits: Tadej Droljc (audiovisual composition), Gregor Zemljič (mastering)

Fail Better

Path: Music – Study Piece \ Fail Better.aif
www: https://soundcloud.com/smrgoda/fail-better
Length: 9’ 28”
Format: Musical composition, stereo audio
Credits: Tadej Droljc (music)

List of Submitted Max Patches

Stringvencer 2.0

Path: Patches \ Stringvencer 2.0 \ Stringvencer_2.0.maxpat

Path to Presentation Video: Presentation Media \ Stringvencer 1.0.mp4

Path to Tutorial Video: Presentation Media \ Stringvencer Tutorial.mov

Accompanying *Max for Live Devices*: NotesToJitMatrix and MultiMap-UDP_Receive

Remarks: needs xray externals to work

Installation of externals: Max->File->Show Package Manager->search for “xray”->install

Application for Optimising the 3D Model for Working Within Max

Path: Patches \ Application for Optimising the 3D Model for Working Within Max \ Application for Optimising the 3D Model for Working Within Max.maxpat

Accompanying Externals: ThinAndIndexForTriangMesh8plane.java

Remarks: needs xray externals to work

Installation of externals: Max->File->Show Package Manager->search for “xray”->install

Connect the Points – Java Version

Path: Applications \ Connecting the Points \ Connect Points – Java Version \ Connect Points - Java Version.maxpat

Connect the Points – JavaScript Version

Path: Applications \ Connecting the Points \ Connect Points – JavaScript Version \ Connect Points - JavaScript Version.maxpat

Connect the Points – Max Version

Path: Applications \ Connecting the Points \ Connect Points – Max Version \ Connect Points - Max Version.maxpat

List of Submitted Max Externals

***connectPointsMultiLimit.java* (Connecting the Points)**

Path: Applications \ Connecting the Points \ Connect Points – Java Version \ connectPointsMultiLimit.java

***ThinAndIndexForTriangMesh8plane.java* (Application for Optimising the 3D Model for Working Within Max)**

Path: Applications \ Application for Optimising the 3D Model for Working Within Max \ ThinAndIndexForTriangMesh8plane.java

List of Submitted Presentation Media

Micro-isomorphism and Tension.mov

Path: Presentation Media \ Micro-isomorphism and Tension.mov

Particle-based Granular Synthesis - Test 1.mp4

Path: Presentation Media \ Particle-based Granular Synthesis - Test 1.mp4

Particle-based Granular Synthesis – Test 2.mp4

Path: Presentation Media \ Particle-based Granular Synthesis - Test 2.mp4

Rhythm of Undulating Geometry.m4a

Path: Presentation Media \ Rhythm of Undulating Geometry.m4a

Rhythm of Undulating Geometry.mov

Path: Presentation Media \ Rhythm of Undulating Geometry.mov

Stringvencer 1.0

Path: Presentation Media \ Stringvencer 1.0.mp4

Stringvencer Tutorial

Path: Presentation Media \ Stringvencer Tutorial.mp4

List of Figures

Acknowledgements

I would like to thank my supervisors Dr Alexander Harker and Professor Pierre Alexandre Tremblay for their extensive support, mind opening guidance and priceless feedback.

My gratitude goes to *Centre for Research in New Music* and *University of Huddersfield* for awarding me a *Denis Smalley Scholarship*. Big thanks also to *Slovenian Ministry of Culture* for *Scholarship for Postgraduate Studies Abroad*.

Many thanks to all teaching staff at the *Centre for Research in New Music* and the *Sound Music Image Collaboration research centre* for inspiring lectures, seminars, colloquiums, talks and presentations. I would like to especially thank Julio d'Escriván, Aaron Cassidy, Bryn Harrison, Monty Adkins, Alexander Harker and Pierre Alexandre Tremblay.

Very special thanks to my postgraduate colleagues and friends for great debates and many unforgettable moments. I would like to especially thank Elías Merino, Jorge Boehringer, Cassandra Miller, Stephen Harvey, Sébastien Lavoie, Oliver Larkin, Frédéric Dufeu, Zeger de Vos and Eleanor Cully.

Massive thanks to my partner Dijana Jež for her love, support, patience and feedback. Special thanks also to my son Teodor and daughter Naja the dog for their love and patience.

1 Research Aims

The core of my aesthetic research is a body of audiovisual work that portrays a quasi-concrete and non-referential audiovisual world. Visual elements are made from digitally generated 3D graphics rendered with lines and points suspended in negative space. The musical side is anchored in post-acousmatic (Adkins et al., 2016) and post-digital (Cascone, 2002) beats-infused aesthetics.

This personal research was inspired by forms and motion characterised by the juxtaposition of organic and artificial qualities, such as cross-sections of natural formations or organic motion in artificially confined spaces. The research was also inspired by the possibilities of the digital world, where such forms and types of motions can be simulated, controlled, transformed, extended beyond the limitations of concrete world and data extracted for audio and visual purposes. Exploration of these features in the digital domain in regard to audiovisual composition represents the essence of this research.

The audiovisual material I work with is structured in a manner to form what I call an audiovisual gestalt, which can be understood as a synergetic audiovisual whole on the level of perception. Sound and image in my gestalts support each other rather than go against one another. There are different ways of creating an audiovisual gestalt and what is common to all of them is what I call a certain degree of spatio-temporal congruency and a certain degree of cognitive complementation between sound and image. These two terms were inspired by two categories of psychophysical binding factors namely structural and cognitive (Spilotros & Parzy, 2010; Tsilionis & Vatakis, 2016; Spence, 2015; Spence et al., 2007).

There were five research aims:

1. To explore the juxtaposition of artificial and organic qualities on various levels of an audiovisual composition, from basic compositional elements to the overall characteristics of the form.

2. To explore a variety of audiovisual structures that result in the formation of an audiovisual gestalt. This variety deals with different ratios of importance between spatio-temporal congruency and cognitive complementation between sound and image.

3. To compose with different hierarchies between sound and image and explore the co-existence of these different hierarchies within a single composition.

4. To explore the audiovisual potential of mapping physics-based geometric motion to sounds and/or images. This includes sound synthesis, manipulation of musical time, visual effects and creation of rhythms.

5. Bearing in mind research aims 3 and 4, to design an audiovisual system that enables me to integrate Cycling 74's *Max/Jitter* patching paradigm for creating real-time *openGL* 3D graphics into the *Ableton Live/Max for Live* DAW paradigm. The system should allow me to work simultaneously with sound and image and combine the flexibility and strengths of these two worlds.

In Chapter 2 the portfolio of works is presented. Chapter 3 outlines the theoretical framework (*isomorphic audiovisual paradigm*) encompassing models of perception from gestalt psychology (isomorphism and grouping principles) and psychophysics (binding factors). Chapter 4 discusses aesthetic influences and the integration of these influences in my audiovisual language. The topics of time, compositional indeterminacy and what I call audiovisual hierarchies, audiovisual tightness and *audiovisual escapology* are all discussed in relation to my work in chapter 5 which deals with composition. Chapter 6 focuses on the workflow and the development of the custom-made audiovisual system that extends my working environment. The system spans across two pieces of software, namely *Max* and *Live*. Chapter 7 presents the solutions to the main technical obstacles encountered in the context of aesthetic research.

2 Portfolio

In this chapter I will present a body of work that form my compositional portfolio. The bulk of portfolio consists of audiovisual works. Some of the them were made in collaboration with Elías Merino where he composed the music and I provided the visuals. We work under the name *Synspecies*. Other audiovisual works are my solo works. Part of portfolio is also my study piece entitled *Fail Better*, which is a musical composition without visuals.

2.1 Capillaries Capillaries

Length: 22’ 42”
Format: Audiovisual composition, UHD 4K, 60FPS, stereo audio
Credits: Tadej Droljc (audiovisual composition), Gregor Zemljič (mastering)

Capillaries Capillaries is an audiovisual composition that does not prioritise any established audiovisual hierarchy. Therefore, the piece does not represent a visualisation of music or music for an image but rather a tangle of audiovisual interactions. In a predominantly bottom-up compositional approach, most of the compositional decisions came as a consequence of cross-fertilisation between both modalities.

The work extends the idea of an audiovisual object to an audio-visual-time object - that is, a tangle of interactions between sound, geometry and time. The audio-visual-time object is treated as a (meta)physical material inside a high-voltage continuum. The tensity of the continuum arises from the imbalance between various forces that pull the material in varied directions. Since the material is a complex audio-visual-time object, the result is often impossible to predict. Exploring the emerging structures and their behaviors is the main focus of the piece.

The concept represents the translation of the id, ego and super-ego model to the audiovisual world. Just as the ego is established between the two groups of conflicting forces that constitute, distort, constrain, liberate, punish or disperse our subjectivity, so too is the material of *Capillaries Capillaries*. What we observe throughout the work is one and the same material transformed in different ways.

The emerging structures were hence approached from two opposite angles. Some parts of the piece are predominantly “id driven” and are based on “hands off” generative techniques where time, rhythm and sound originate from various methods of geometry sonification - hence the unexpected bursts of energy, jerky rhythms and random time.

Other parts on the contrary reflect the super-ego like internalisation of cultural norms - hence the metric re-organisation of audiovisual material into tidy sound-geometry-time gestalts or comprehensible shapes. However, the main focus of the piece is the space in between where the forces of id are tamed and the rules of super-ego bended - where noise becomes form and form becomes noise.

2.2 Singing Sand

Length: 12’ 34”
Format: Audiovisual composition, UHD 4K, 60FPS, stereo audio
Credits: Tadej Droljc (audiovisual composition), Gregor Zemljič (mastering)

Singing Sand was inspired by the sonic potential of abstract 3D computer graphics. The central element of the composition is a physics-based visual material whose movement is sonified in real-time. The piece explores what various particle fluctuations, tensions or shape-morphings sound like, what internal rhythms they create and how that affects our perception of the visuals in return.

Sand resembling visual particles are sonified by mapping their individual velocities to various parameters of individual grains inside a custom-made granular synthesizer. At the same time the velocities determine the colours of the particles which together creates a spectrum of colours and sounds. This results in perceptually tight audiovisual bonds, which is a crucial element of this piece as the composition focuses on the gestural qualities of the audiovisual material.

The aim of the composition was to explore how such audiovisual material could function in a free audiovisual paradigm and also how to meaningfully force it onto the grid of harmony and metric rhythm. In the latter case the material created liquid grooves on top of dub-tech influenced fixed rhythmic elements, as well as ever changing spectral swirls that emerged from cross-breeding an originally noisy and non-stable sound source with pitch-based material.

The piece represents an attempt to compose from within an audiovisual paradigm where most of the compositional decisions came as a consequence of cross-fertilisation between both modalities.

2.3 Synspecies – Teaser (Promo Video)

Length: 1’ 19”
Format: Audiovisual composition, UHD 4K, 60FPS, stereo audio
Credits: Elías Merino (music), Tadej Droljc (visuals), Gregor Zemljič (mastering)

Synspecies is an audiovisual project created by Elías Merino and Tadej Droljc. It is inspired by virtual ecologies – unstable morphological spaces that emerge from the interaction of co-existing unrelated entities with the void.

Synspecies are multi-scaled audiovisual objects that are cross pollinating, mutating, adopting and fighting for their space and existence. They co-create a world of unreal physicality that is constantly kneaded by violent forces. Their existence appears to us as fractured narratives in the perceptual phenomena that resemble a chain of synchronised chaos.

Synspecies feed on tangles made by computer-based code and analogue signals. They originate from Stockholm’s EMS *Buchl*a and *Serge* modular systems, algorithmic composition and 3D geometry synthesis. The vivid abstract architectures of *Synspecies* fluctuate in an ambiguous pulsating time, hybridising and shaping new spaces confined in a perpetual disintegration process.

This video represents my first attempt to use my audiovisual system for purely music visualization purposes.

2.4 Synspecies – Spaceless Latitudes

Length: 6’ 28”
Format: Audiovisual composition / Installation, 3 x HD, 60FPS, 4.1 audio (installation version); UHD 4K, 60FPS, stereo audio (portfolio version)
Credits: Elías Merino (music), Tadej Droljc (visuals), Gregor Zemljič (mastering)

Spaceless Latitudes is a 3 video channels and 4.1 surround sound audiovisual installation. It is a fixed media installation based on a video loop. The composition consists of 4 contrasting and self-standing sections that are not interrelated.

The piece represents my first exploration of multi-channel video (triptych). The compositional process was to first compose the music and later add the visuals. However, during the composition of both elements, music and visuals, Elías and I were in constant dialogue, so Elías was composing music by visualizing the audio gestures and overall progression of the piece in his mind. An important element of our conversations was the adaptation of the spatio-temporal features of audio articulations to the rate that could be meaningfully matched with visuals in a micro-isomorphic manner (see section 3.3.1 for micro-isomorphism).

The installation *Spaceless Latitudes* portrays the first work of the *Synspecies* project.

2.5 LPM Promo Video

Length: 0’ 58”
Format: Audiovisual composition, UHD 4K, 30FPS, stereo audio
Credits: Tadej Droljc (audiovisual composition), Gregor Zemljič (mastering)

LPM Promo Video was made for the Rome 2018 edition of the *LPM festival*. The request of the festival was that it needs to be performable, hence real-time, and that it needs to match the style of their visual design.

For this short piece I used an unfinished piece of music I composed with the help of *Stringvencer*, which is a tool that treats timeline as a deformable string (see section 5.3.1 for *Stringvencer*). I adopted the piece in order to fit better into the context of the festival, which combines anything from experimental audiovisual art to VJ-ing on beat based music. The reason this piece is included in my composition portfolio is that it shows the use of the *Stringvencer* tool in a more club oriented music. The colourisation of the particles on the other hand was done in the same manner as with *Singing Sand*, that is by mapping the velocity of the particles to colour.

2.6 Fail Better

Length: 9’ 28”
Format: Musical composition, stereo audio
Credits: Tadej Droljc (music)

Fail Better is a study piece in which I was developing my sound world and testing the first prototype of *Stringvencer* (see section 5.3.1 for *Stringvencer*). The overall groove of the piece was done using a *Stringvencer* as a slight time distortion tool and hence a groove tool. At the very end of the piece the *Stringvencer* was used to radically deform time - to an extent of restructuring the musical material. This was done by dragging around the string-based timeline as it can be seen in the excerpt (1:08 – 1:25) from the video *Stringvencer 1.0* (Droljc, 2015).

This piece influenced the development of my musical language in two ways:

1. As I was dragging and throwing around the pre-composed loop in a *Stringvencer* timeline, the density of sound events was very high. Such use of *Stringvencer* resulted in perceived sonic gestures rather than perceived time distortion. Since the *Stringvencer* was originally triggering drum samples stored in *Live* samplers, these audio gestures became quickly too similar sounding and hence annoying to work with. Changing the samples into 'noisy' and 'glitchy' sounds helped, as it is harder to memorise them. These 'noisy' and 'glitchy' sounds became part of my sound world.

2. I realised I prefer to use *Stringvencer* as a tool for creating slow temporal deformations rather than fast. *Stringvencer* was however initially designed for fast temporal deformations.

3 The Isomorphic Audiovisual Paradigm

3.1 Introduction

3.1.1 Outline of the Isomorphic Audiovisual Paradigm

The aim of this chapter is to present a model of audiovisual relationships that I have developed for thinking and talking about my work. I call my model an isomorphic audiovisual paradigm, or shortly isomorphic paradigm. Isomorphic, in its Greek origin means sameness of form, in my case primarily refers to similarity between temporal and cognitive structures of audio and visuals. My model is an adaptation and fusion of several existing concepts of isomorphic relationships from non-referential audiovisual art (Coulter, 2010), film (Audissino, 2017) and psychology (Luchins & Luchins, 2010) to mathematics (Palmer, 1999).

The central idea behind the isomorphic paradigm is a spectrum of isomorphic audiovisual relationships. On one end of the spectrum we find audiovisual objects with a high degree of spatio-temporal congruence, where sound and image appear as coming from one and the same source while on the other end of the spectrum we find loose audiovisual clusters with medium degree of spatio-temporal congruence and high degree of cognitive complementation. In latter case sound and image share some common qualities and hence also belong together, but the two mediums do not appear as emanating from a single source. Audiovisual objects and audiovisual clusters are two key types of cross-modal gestalts that I use in my compositions.

An example of an audiovisual object could come from any congruent spatio-temporal morphology between sound and image, such as the video recording of a speaking mouth - to borrow an example from a referential world. An example from a non-referential audiovisual world might be a sphere, whose amount of shape deformation is temporally congruent with the amount of a sine tone frequency modulation. On the other hand, examples of audiovisual clusters could include a noisy visual texture accompanied by a noisy drone (cross-modal correspondence between the noisiness of sound and image[1]), or a film soundtrack that follows the emotional fluctuations of the narrative onscreen. An audiovisual cluster hence does not form an object but gives a listener-viewer a sense of

[1] The cross-modal correspondences I use might be culturally and personally conditioned.

belonging together due to congruent expressive and affective qualities between sound and image.

In between these two types of audiovisual gestalts, or rather in-between the two types of underlying relationships between sound and image lies a spectrum of possibilities that can be exploited for achieving cross-modal contrast inside an audiovisual composition. The isomorphic paradigm hence does not rely on the compositional contrast which could be achieved between parallel and contrapuntal audiovisual relationships, using the classical film analysis terms (Birtwistle, 2017; Monaco & Lindroth, 2000; Burt, 1994; Buhler & Neumeyer, 2000), but rather exploits the variety of tight and loose isomorphic relationships.

3.1.2 The Isomorphic Paradigm in a Wider Context

In 'Analysing Musical Multimedia' Nicholas Cook proposed three universal models of audiovisual relationships (Cook, 1998). The three models are *conformance, complementation* and *contest.*

Audiovisual relationships that I would categorise as isomorphic, are, in the literature, often compared to Cook's *conformant* audiovisual relationships (Audissino, 2017; Grierson, 2005; Auslander, 2013; Kerins, 2010). *Conformance* describes situations in which sound and image communicate the same message. At the other extreme is *contest* that refers to contradicting messages of sound and image. In between is *complementation*, where the two messages complement each other.

Analysing my work using Cook's models, the audiovisual relationships could be categorised as *conformant* or *complementary.* An example of *conformance*, relevant in the context of this research, would be a very simple object with one-to-one parametric mapping where louder sine wave would correspond with a bigger circle. What makes such relationship *conformant*, according to Cook's model, is not the fact that parameters are directly mapped but the similarity between sound and image in terms of their meaning. The relation bigger-bigger clearly communicate the same message through both mediums. Whist such simple examples exist in my work there are also more complex examples where an image might completely transform the meaning of audio. In such cases the audiovisual correspondences are not so explicit - certain qualities of sound could for instance correspond to a number of visual features while certain features might not directly correspond with anything. Such scenario would offer an exchange of semantic attributes between sound and image (see *metaphore* model for details on exchange of semantic attributes: Cook, 1998) which would provide a listener-viewer with a *complementary* meaning.

The reason why I am not using either *conformance* nor *complementation* models in my theoretical framework is that Cook's categorisation is based on the messages that sound and image are communicating[2]. I find these models insufficient for my particular needs, as my requirement is to categorise audiovisual relationships based on the types of audiovisual gestalts they create. My proposed model hence tries to categorise audiovisual material not on the basis of its semantic attributes but rather attempts to establish a phenomenological differentiation between objects and clusters.

3.2 Theoretical Foundations

In this subchapter I will present a theoretical basis necessary for the understanding of terms I use in this text. The concepts and models presented originate mainly from psychophysics and gestalt psychology.

3.2.1 Binding Factors

> *In many everyday situations, our senses are bombarded by many different unisensory signals at any given time. To gain the most veridical, and least variable, estimate of environmental stimuli/properties, we need to combine the individual noisy unisensory perceptual estimates that refer to the same object, while keeping those estimates belonging to different objects or events separate. How, though, does the brain "know" which stimuli to combine?*
>
> (Chen, 2016).

This question is known as the binding problem and the part of the binding problem that is most relevant for this research deals with cross-modal binding. Cross-modal binding factors "map onto a number of Gestalt grouping principles that have been put forward

[2] Cook is applying his *conformance, complementation* and *contest* models also to non-referential audiovisual materials. An example in his book *Analysing Musical Multimedia* would be his critique of the use of the light instrument in Skriabin's *Prometheus*: »the *luce* part literally *does* add little...the faster part simply duplicates information that is already present in the music« (Cook, 1998, p. 40). On page 100, where he developes his model of *conformance*, he refers to this example.

previously to describe how sensory information is perceptually organised within individual sensory modalities (i.e., intramodal perceptual grouping or stream segregation)" (Spence et al., 2007).

Psychophysical binding factors are examined through experiments using simple audiovisual stimuli. In fact, in most cases "experimental design is created to make sure that only one independent factor … is varied and that the results … can be measured and quantified" (Langkjaer, 2015, p. 4). The audiovisual material I work with is, on the other hand, very complex and some of the phenomena I observe and exploit in my works cannot be directly linked to the examples from the literature. When composing with audiovisual material for example, we need to be also aware that spatio-temporal structures, such as gestures, rhythms and patterns, can possess cognitive qualities – they can induce associations and hence place spatio-temporal events into a wider network of meaning.

Following the terminology established in psychophysics, the binding factors are divided into two categories: spatio-temporal and cognitive (Spilotros & Parzy, 2010; Tsilionis & Vatakis, 2016; Spence, 2015; Spence et al., 2007)[3].

3.2.1.1 Spatio-Temporal Binding Factors

In this section four spatio-temporal binding factors will be presented. The definitions will be adopted according to the needs of my compositional model.

Temporal Proximity

Temporal proximity refers to the perceptual congruence of temporal structures between sound and geometry. In certain cases, with very explicit and simple correspondences, temporal proximity can be reduced to the purely physical and hence measurable dimension of time. On the other hand, when a more complex audiovisual activity with longer durations takes place, temporal proximity becomes a phenomenological factor – it refers to the perceived temporal synchronisation of either micro-morphology between sound and image (in case of audiovisual objects) and on the other hand to the perceived temporal alignment of the macro-structural elements within sound-image relationships (in case of audiovisual clusters).

[3] Spatio-temporal binding factors are in literature often refferd to as structural.

Spatial Proximity

Spatial proximity refers to the degree of spatial congruence between the position of the visual gestalt and the perceived position of the auditory stream. Whilst spatial proximity affects the cross-modal integration, this factor is not as important as temporal proximity. This is due to the spatial ventriloquism effect (Slutsky & Recanzone, 2001), in which the presence of a visual object captures and modulates the perceived spatial properties of an auditory stream. Hence, for instance, one can use a very wide stereo sound in combination with a central visual gestalt and yet the cross-modal gestalt would still form centrally.

Spatial proximity is an important factor in case of audiovisual objects while in case of audiovisual clusters it does not play a significant role.

Time Windows

A time window refers to the time span in which we experience a certain audiovisual gestalt in a stable state. Such time window are created either with an edit[4] or a step-transformation of audio and/or visual properties. Shorter time windows affect the quality of the cross-modal integration as the listener-viewer does not have enough time to comprehend the audiovisual gestalt. I use shorter time windows to induce a sense of unpleasantness, confusion or tension in the listener-viewer (see *Spaceless Latitudes* 3:48 – 4:00). According to my experiences, the exact duration at which this happens depends on factors such as visual complexity of an audiovisual object and its offset from the centre[5]. However, a rough estimation of duration for a short time window would be around 0.5 second. I also use relatively short successive time windows, but generally longer than 0.5 seconds, to increase the perceived visual energy (see *Singing Sand* 9:53 – 10:17). On the other hand, I use extremely short time windows (up to few hundred milliseconds) to generate simple audiovisual objects on the basis of *synchresis*[6] - audio transients combined with short bursts of visual energy create a sense of an audiovisual object that however does not display any morphology (see *Capillaries Capillaries* 9:53 – 10:17).

[4] What appears as *video edit* does not necessary refer to post production technique of editing – see section 7.2.2 for more details.
[5] According to the research by Talsma et al., centrally presented visual gestalts are integrated faster as peripherally presented gestalts (Talsma et al., 2007). This research paper, however, does not discuss psychological factors such as tension.
[6] Chion's term *synchresis* encompasses precisely synchronised audiovisual events as well as audiovisual phenomena which I refer to as intramodal peceptual grouping (see 3.2.1.1), where the synchronisation between the two mediums is phenomenological. In this thesis I am using the term *synchresis* to refer to the temporal synchronisation of short and isolated audiovisual events.

Intramodal Perceptual Grouping

Intramodal perceptual grouping is a phenomenon in which perceivable patterns in one modality affect the perceived organisation of a cross-modal scene. An anecdotal example would be "the flickering lights on a Christmas tree …[that] appear to flash in time with the rhythm of the music" (Spence et al., 2007, p. 520). In this example "the perceptual organization of stimuli taking place in one sensory modality (audition in this case) may be used by the nervous system to help organize (or structure) the information that is simultaneously being processed by other sensory modalities (in this case vision)" (Spence et al., 2007, p. 520). The phenomenon that I exploit in my works does not have to rely on patterns. My observation is that the stronger perceptual organisation in general overrides the weaker one. In this example (*Capillaries Capillaries* 16:02 – 16:15) I believe that musical micro-rhythm dictates the perceived rhythm of visual connections, whilst in this case (*Spaceless Latitudes* 0:40 – 0:55, right screen) the result depends on the focus as the scene is more complex – one possible interpretation is that the visual pattern and visual rhythms suggest the segregation of the auditory stream and interfere with musical micro-rhythms.

3.2.1.2 Cognitive Binding Factors

In this section four cognitive binding factors will be presented. The definitions will be adopted according to the needs of my compositional model.

Cross-modal Correspondences

> *Cross-modal correspondences have been defined as compatibility effects between attributes, or dimensions, of stimuli … in different sensory modalities.*
>
> (Spence, 2015, p. 648).

Cross-modal correspondences "typically constitute relative, rather than absolute, mappings between modality-specific dimensions" (Chen at al., 2017), such as louder (bigger) sound – bigger object, higher visual energy – higher musical energy or sharper timbre – sharper visual texture. These examples rely on concepts that are transferable across both modalities.

Semantic Congruence

> *Semantically congruent stimuli refer to those features/attributes mapped to a common object or category.*
>
> (Chen at al., 2017)

As opposed to the relative mappings encountered in the case of cross-modal correspondences, semantic congruence refers to absolute mappings, such as *image of a barking dog – sound of a barking dog*. In a non-referential audiovisual world semantic qualities are rare. However, there are examples in my work where certain sonic and visual features point towards a single concrete element. Examples of such audiovisual pairs are: *sound with 'electrical qualities' – motion resembles the response to an electric shock* (*Capillaries Capillaries* 4:40 – 4:47) or a *visual object resembling a rock – sound with a rock -like timbre* (*Capillaries Capillaries* 11:40 – 11:43). However, as discussed in the paper 'Assessing the Role of the 'Unity Assumption' on Multisensory Integration' (Chen at al., 2017), there is an overlap between semantic congruence, cross-modal correspondences and the unity assumption. Thus, examples of semantic congruence in my work tend to also involve cross-modal correspondences.

The Unity Assumption

> *The 'unity assumption' is an observer's assumption, or belief, that two or more unisensory cues belong together (i.e., that they come from the same object or event…)…The use of the term 'assumption' or 'belief' might lead one to suggest that this modulatory factor relies on the observer being consciously aware of the connection, or relation, between the component unisensory signals… [However] the unity assumption need not be represented consciously.*
>
> (Chen at al., 2017)

The Unity assumption in my work arises mainly as a consequence of tight audiovisual integration, by utilising other cross-modal binding factors, and is heavily associated with the spatial and temporal ventriloquism effect (see *Capillaries Capillaries* 2:40 – 3:42)

Visual and Audiovisual Complexity

> *It has been found…that auditory stimuli are known to capture attention easily (Schroger and others 2000) and also that the processing of auditory stimulus features occurs generally faster than that of visual stimuli (Woldorff and others 1991, 1993).*
>
> (Talsma et al., 2007)

The cited text links to my observations that we need more time to process complex visual objects as compared to complex auditory streams. When I work with shorter time windows this complexity needs to be considered. Audiovisual complexity, in which both audio and visual elements are complex, leads to a loss of perceived correspondence between sound and image. The reason for this is that there are too many different elements for the mind to group together. In that sense the audiovisual complexity encourages the listener-viewer to create their own cross-modal connections from elements that might not have any common structural basis. This relates to the concept of intramodal perceptual grouping (see *Spaceless Latitudes* 0:40 – 0:55).

3.2.2 Isomorphism

> *The term isomorphism literally means equality or sameness (iso) of form (morphism). In mathematics an isomorphism between two systems requires a one-to-one correspondence between their elements …, which also preserves structures…In Gestalt psychology, the one-to-one correspondence between elements is not required; similarity of structures is required.*
>
> (Luchins & Luchins, 1999, p. 208)

In Gestalt theory, isomorphism resides in mental activity - "a correspondence between the configurations of the external objects as we perceive them and the configurations of the energy field in the brain structures responsible for the perception of those objects" (Audissino, 2017, p. 102). According to Rudolf Arnheim, isomorphism was introduced "to describe similarity of structure in materially disparate media" (Arnheim, 1986, p. 13) and to provide a psychophysical explanation of why, for example, "a dance and a piece of music accompanying it can be experienced as having a similar structure… [and why] such structural kinship …[can be] directly and spontaneously experienced" (Arnheim, 1986, p. 13). Hence, we can think of isomorphism as a "correspondence between Gestalt qualities from one field of experience and another" (Audissino, 2017, p. 106) or in this context between the audio and visual modalities.

In this portfolio, isomorphic relationships are based on two fundamental criteria: a certain degree of proximity between spatio-temporal structures of sound and image and a certain degree of cognitive complementation between sound and image. Therefore, my idea of the isomorphic paradigm draws both on the notion that isomorphism can be understood in terms of searching for common gestalt qualities across various media, and also on a structural approach to analysis of audiovisual phenomena based on binding factors taken from psychophysics. The first of these represents a more holistic and intuitive approach to discussing audiovisual relationships. The second allows a focussed analytical approach to individual components in an audiovisual scene. My idea of isomorphic paradigm will be developed in the following sub-chapter 3.3.

3.2.3 Cross-Modal Gestalts

Whilst there is some disagreement in literature on the definition of the term cross-modal gestalt (Spence, 2015), broadly speaking it can be understood as a percept whose elements derive from different modalities, but appear as a single whole.

Various authors use the term 'cross-modal gestalt' to describe percepts based on multisensory organisation (Spence, 2015). According to Spence (for whom the general use of the term cross-modal gestalts is controversial), cross-modal gestalts can be considered as cross-modal correspondences when understood through the prism of Gestalt grouping by similarity (Spence, 2015). This means that groups may be formed from corresponding elements across different modalities, thus forming a gestalt.

An important part of cross-modal gestalt is also what Chion would call an *added value* which can be understood as a synergetic effect when adding together sound and image. The way I perceive the *added value* is through gestalt principle of reification (see Figure *1* for the example of reification principle). "Reification is the constructive or generative aspect of perception, by which the experienced percept contains more explicit spatial information than the sensory stimulus on which it is based" (Schreuder, 2014, p. 642).

Figure 1: “Idesawa’s Spiky Sphere” (Lehar, 2008), illustrative example of Gestalt principle of reification.

In Figure 1 the configuration of cones outlines a sphere which is not actually present in the image. In this case, "the whole is *something else* than the sum of its parts" (Kofka, 1999, p. 176). In a similar way, configurations of sonic and visual elements can be used to create new meanings or added values. An example can be seen and heard in *Synspecies – Teaser* (0:56 – 1:08).

3.3 The Isomorphic Paradigm: From Micro to Macro

In this subchapter two types of audiovisual relationships that underpin two different types of audiovisual gestalts will be defined: micro-isomorphic audiovisual relationships that underpin audiovisual objects and macro-isomorphic audiovisual relationships that underpin audiovisual clusters.

Micro-isomorphic audiovisual relationships are attributes of audiovisual objects. They demand a precise spatio-temporal congruence between sound and image, with an emphasis on temporal congruence, while cognitive complementation is not a constituent factor. Macro-isomorphic audiovisual relationships in audiovisual clusters on the contrary allow looser spatio-temporal congruence between sound and image whilst demanding a higher degree of cognitive complementation.

Micro-isomorphic and macro-isomorphic relationships hence determine two different configurations of binding factors.

3.3.1 Micro-isomorphism and the Audiovisual Object

In this thesis the audiovisual object is a discrete mental structure (a type of audiovisual gestalt), based on the spontaneous belief, that sound and image come from a single object. The image, which in my opinion has a stronger semantic value, has to appear as an object, whilst the sound may consist of one or more auditory streams. The majority of auditory streams however need to correspond with the elements of the visual object – they need to have congruent spatio-temporal structures on a micro level. Spatio-temporal congruency on a micro level is the essence of micro-isomorphic relationships. The audiovisual disparities need be less than approximately 50 – 100 ms in order to be perceived as simultaneous (Slutsky & Recanzone, 2001). Cognitive complementation, which is not a constituent factor, however, enhances the audiovisual integration into an object.

The bindings of audiovisual objects can span between tight and loose.

Tight audiovisual objects need to have a very explicit audiovisual morphology or very explicit rhythm in order for the congruence of spatio-temporal structures to be clearly perceivable. A large majority of visual elements and auditory streams need to be corresponding.

Loose audiovisual objects have less visual elements corresponding with the auditory streams. This could be the consequence of certain visual elements and auditory streams actually not being mapped together, or alternatively it could be the consequence of an audiovisual complexity where all the streams and elements are mapped but due to overwhelming amount of information one cannot identify the matching correspondences.

An example of a loose audiovisual object that gradually becomes tighter can be observed in *Capillaries Capillaries* between 4:07 and 4:58. The excerpt starts with one visual object, consisting of three distinct elements: 1) a larger portion of an object that is mostly hidden in the dark, 2) brighter smaller portion consisting of larger dots that dynamically connect with lines and finally 3) flashing lights. Similarly, there are three audio streams: 1) stream of slowly interweaving sounds, 2) sound of larger dots being connected with the lines and finally 3) the sound of flashing lights. The last two visual elements are in micro-isomorphic relationship with the last two auditory streams. Their explicit rhythmic patterns are synchronised while the quality of sound and image also cognitively complement each. If there would be only these two pairs corresponding, the audiovisual object would be tight. But since there is a third pair, the larger portion of the visual object and accompanying stream of interweaving sounds, whose temporal structure is only occasional congruent, the overall audiovisual object is loose. Towards 4:47, before the scene changes into white, the object becomes tighter and tighter as there is another micro-isomorphic layer added: shaking of the visual object synchronised with rough electric sounds. Also, the micro-isomorphic pair of connecting lines becomes denser so the micro-isomorphic relationships start to dominate the scene. However, as we enter the white sequence, the object becomes extremely tight. With the exception of few reversed kick drums, sound and image are entirely in micro-isomorphic relationship.

The formation of the audio-visual object in cases where not all the audiovisual elements are corresponding I believe follows the Gestalt *law of Prägnanz*, which states that the gestalt formations tend toward maximal simplicity "or toward the formation of 'good form'" (Asch, 1968, pp. 159-160).

In referential world, an audiovisual object can be formed purely on the basis of *semantic congruence* – an example would be a car that stands still with its engine running. However, in non-referential audiovisual world, some explicit gesturality or rhythmic elements are needed. "Behavioral findings have shown that the near-simultaneous presentation of visual and auditory stimuli and their having common locations are 2 fundamental properties that facilitate the integration of audiovisual stimuli into a multi-sensory percept or multisensory object (Lewald and Guski 2003)" (Talsma et al., 2006). For this reason, my computer generated quasi-concrete audiovisual object shares some similarity with concrete reality.

3.3.2 Macro-isomorphism and the Audiovisual Cluster

The audiovisual cluster is a mental structure, a type of audiovisual gestalt, based on the spontaneous belief, that sound and image complement each other into a larger single whole. This larger audiovisual whole I call an audiovisual cluster. Despite that sound and image fit together, they do not appear as coming from the same source, hence they do not form an object. The sense of audiovisual complementation is achieved through the strong cognitive complementation and occasional congruence of spatio-temporal structures on a macro level. Such configuration of binding factors will be called macro-isomorphic.

The idea of macro-isomorphic relationships was inspired by Emilio Audissino, who refers to macro-isomorphism only as isomorphism. [Macro-]Isomorphism for him can be observed when comparing "gestalt qualities of the music, which can be roughly identified with what Meyer calls 'secondary parameters (Meyer, 1996: 209)'" (Audissino, 2017, p.108), with expressive and affective parameters of the image. I refer to all these parameters as macro parameters, or macro gestures. The fluctuation of macro parameters creates a macro spatio-temporal structure.

As with audiovisual objects, audiovisual clusters can be tight or loose.

An example of what I would classify as a tight audiovisual cluster can be found in a book *Film/Music Analysis* by Emilio Audissino:

> *… think of the sequence in Hook (Spielberg, 1991) in which Peter Banning retrieves his 'happy thought' and thus is able to fly again and reverse to his previous Peter Pan status. As soon as the happy thought is secured, the music builds up, creating a crescendo in volume and orchestral texture, with flutes and piccolo bubbling in the higher register to express the excitement that is hardly containable and about to explode; then, when an exhilarated Peter takes off to fly over Neverland, the music explodes in a tutti fortissimo exactly as Peter pops out like a missile into the sky;*
>
> (Audissino, 2017, p.111)

What makes this example a tight cluster, is not just a very strong cognitive complementation, but also an occasional congruence of macro temporal structures. As shown by Scott Lipscomb in his research entitled *Cross-Modal integration: Synchronization of auditory and visual components in simple and complex media*, the congruence of macro temporal structures does not need to be precise in order to create a convincing audiovisual integration (Lipscomb, 1999).

The temporal structure of audiovisual clusters hence derives from macro gestures, that function as temporal anchors. The audiovisual material that exists in-between them, can nevertheless possess a large degree of temporal freedom between sound and image. The temporal structure of audiovisual clusters is hence more or less personal – depending on how emphasised and frequent the macro gestures are. When macro gestures are subtler and less frequent, the mind can easily start to form its own connections and hence its own temporal structures. That would be an example of a loose audiovisual cluster (see *Capillaries Capillaries* 16:18 – 17:25).

3.3.3 The Compositional Function of Micro-isomorphism

In my works I use micro-isomorphic relationships predominantly to emphasise tension.

In non-referential audiovisual world, according to my experience, sonic gestures create a stronger feeling of tension than visuals. Therefore, an essential factor of audiovisual tension is musical tension. Musical tension is in my works predominantly amplified through micro-isomorphic visual gestures that emphasise the musical gesturally or musical rhythms. This could be explained through selective attention whose key role is "to serve the purpose of selectively enhancing perception (Hopfinger and Mangun 1998)" (Talsma, 2006, p.679). Hence tighter the audiovisual object the more emphasised the audiovisual tension.

The effect of micro-isomorphism in relation to tension is demonstrated in a video example *Micro-isomorphism and Tension*. When spatio-temporal structures between sound and image are congruent, the tension is stronger as when spatio-temporal structures are incongruent (in the latter case the audiovisual object falls apart).

Beside amplifying the tension, the compositional function of micro-isomorphism in my works could be to change the micro structure of music (or visuals) through the principle of intramodal perceptual grouping.

3.3.4 The Compositional Functions of Macro-isomorphism

In my works I use macro-isomorphic relationships predominantly to emphasise release.

As with micro-isomorphism, releasing the tension in my works is predominantly conditioned by music. Hence, in order to emphasise the release with macro-isomorphic visuals the music needs to deliver a sense of release. Visual amplification is then focusing on the rendition of musical gestalt qualities that give us a sense of release. Selective attention should be in this case targeting macro audiovisual qualities.

According to my observations the release qualities of macro-isomorphic relationships depend on the degree of temporal freedom between synchronised macro gestures. Synchronised macro gestures function as a push effect that pushes the listener-viewer temporarily into a liberating space of purely cognitive complementation. Heavily emphasised (and synchronised) macro gestures hence allow a higher degree of temporal freedom in between them.

One way to additionally emphasise and synchronise macro gestures is by using *montage* to create edits at the peak of gestures – and/or on the beats in case of beat based music. This has been exploited in numerous music videos (Stern, 1996; Goffey & Hawley, 2005; Shadforth, 1999). From a macro-isomorphic perspective editing on the beat in music videos solidifies the audiovisual macro temporal structure. However, it is also important to consider the groove of the music when amplifying the release qualities of music as the release is a function of push and pull features of the groove. Edits that fall 'near the beat' and then at the moment of down-beat 'on the beat' can reflect the push and pull effect of the musical groove and hence amplify the release (see *Capillaries Capillaries* 16:18 – 17:25).

3.3.5 The Spectrum of Isomorphic Audiovisual Relationships

The spectrum of audiovisual relationships span between a tight audiovisual objects and loose audiovisual clusters.

There are many possible trajectories of how to move between the micro-isomorphic and macro-isomorphic typology. An example could be a tightly synchronised audiovisual gesture (micro-isomorphism), that gets stretched in time from 'one second' to 'one minute'. The result would be a slowly evolving time-stretched drone accompanied by a slowly morphing visual object. Offsetting the sound and the image for more than 100 ms seconds, which is the maximum offset that would still produce the illusion of temporal ventriloquism (Slutsky & Recanzone, 2001), our experience of the audiovisual cluster would not collapse due to strong cognitive complementation.

Similarly, the effect of micro-isomorphism can be lost with the increasing complexity of audiovisual structures. In the same way that a musical rhythm becomes a texture when too dense, attempts to create dense micro-isomorphism can result in macro-isomorphism. An example would be a very complex audiovisual object, that consists of numerous micro audiovisual gestures, which would be in isolation perceived as micro-isomorphic. However, due to the overwhelming density of micro gestures, the audiovisual scene would appear to us as macro-isomorphic (audiovisual cluster).

In practice, micro-isomorphic and macro-isomorphic relationships can be often observed simultaneously in my audiovisual material. Various combination of the two can produce gestalts that are sometimes difficult to place on the isomorphic spectrum.

4 Towards Composition: Influences and Aesthetics

In this chapter I will focus on examples from audiovisual art that had an influence on the development of my audiovisual language. Specific audiovisual objects and phrases from various artist will be exposed and mapped to the compositional units or techniques that I use in my work. In order to do that, I will observe them through the prism of micro-isomorphic and macro-isomorphic model presented in sub-chapter 3.3.

4.1 Objects with Static Audiovisual Relationships

In my work, static audiovisual relationships refer to stable perceptual correspondences between auditory streams and visual elements. This can be achieved either through manual synchronisation of audio and visual perceptual parameters; one-on-one mapping on the technical level; or combination of the two.

For static audiovisual relationships, I mostly use real-time one-to-one mappings as this allows me to create objects with multiple complex interactions that I can interact with easily. My influences come from two different sources. One source would be actual audiovisual objects, fully generated in analogue or digital domain, while another source would be human performers using non-conventional body-computer interfaces to control sound with the expressive qualities of their bodies – which I perceive also as audiovisual objects. A clear example of the first of these influences can be found in John Coulter's work *Abide With Me* (approximately between 4:05 and 4:45 (Coulter, 2009)). In this work Coulter mapped the amplitude of an oscillator to the brightness/size of the accompanying dot. Additionally, he mapped the frequency of an oscillator to the 'y' coordinate of the dot. An example of the latter influence that includes body-computer interfaces would be *BioMuse* project from Atau Tanaka, where he is using hand gestures to control amplitude and pitch of the sound source (Tanaka, 2008). Merging the two influences together resulted in my first particle-based granulator (Droljc, 2014), where the idea was to extend the expressive limitations of the human body with a physical simulation of particles. In this first version I mapped the 'y' position of the particles to the pitch of individual grains. In the second version I mapped the velocity of the particles to the pitch, sample reading position, gain and cross-fade between two source samples for each individual grain (Droljc, 2017). See section 5.1.1 for more information on mapping. The latter implementation was used in the piece *Singing Sand* (0:00 – 2:30)

One-to-one parametric mapping, however, does not guarantee a convincing correspondence. Achieving a convincing correspondence between sound and image is context specific and at least in my cases, often involves non-linear mapping, slew-limiting and mapping a parameter from one modality to numerous parameters in the other modality. All these elements are part of the last example.

4.2 Objects with Dynamic Audiovisual Relationships

My use of audiovisual objects with dynamic audiovisual relationships are based on constantly changing audiovisual pairings of visual elements and auditory streams. This means that over time one visual element can correspond to different auditory streams, hence, such objects appear inconsistent and unpredictable.

An influential example comes from Ryoichi Kurokawa's composition *syn mod.2* (Kurokawa, 2014). Throughout this piece, Kurokawa is constantly constructing and deconstructing the audiovisual object cross-modally by establishing very precise relationships between sound and image, only to then change the corresponding audiovisual pairs. Using a combination of dynamic audiovisual relationships and *synchresis* (Chion, 1994), he breaks apart the consistency of the constructed audiovisual reality with a salvo of perplexing sounds and images, and after the very brief "shock therapy" is over, the audiovisual object is re-formed on a basis of different audiovisual relationships. The new audiovisual object, that listener-viewer essentially perceives as the same object, consists of the same sonic elements, same visual elements, but different audiovisual relationships.

I use this influence in my works in two different ways: either to make audiovisual objects more interesting and unpredictable – in such case the rate of change would be much slower as in case of Ryoichi Kurokawa[7] - or on the other hand to create perplexing effects which I prefer to be less saturated with informations as in works of Kurokawa. The first example can be observed in *Capillaries Capillaries* (2:47 – 3:40) and is discussed in detail in section 5.3.7 whilst the latter example can be observed in the video *Synspecies - Teaser* (00:38 – 00:47).

[7] Dynamic audiovisual relationships are not very obvious in works of Ryoichi Kurokawa, due to a high rate of change, information saturation and also simultaneous presence of various macro-isomorphic elements. Such combinations create very perplexing audiovisual objects.

4.3 A Visual Object that "Dances" to the Music

In my works, some visual objects can "dance" to the music through gestures that are synchronised with the most emphasised sounds in the audio stream – sounds with higher structural hierarchy. This can form either micro-isomorphic or macro-isomorphic audiovisual relationships or a combination of the two. I use this idea in a very abstract sense - a "dancing" visual object can be a physics or pseudo-physics-based visual object that responds to a sonic excitation. Reoccurring synchronisation of emphasized sounds with visual excitation and hence visual gesturality creates a sense of micro-isomorphism while the free-floating physics "in-between the beats" form a macro-isomorphism with the rest of the music. Examples would include an excerpt (3:14 to 3:30) from piece *Seething* from Max Cooper & Andy Lomas (Cooper & Lomas, 2014), *Lissajous #2* from Jedrzej Borowski (Borowski, 2015*,* or *SuperCollider Shape* from Joao Martinho Moura (Moura, 2011).

The implementation of the sonic excitations in my work is achieved via any of three approaches described in the first paragraph of sub-chapter 4.1: manual synchronisation of audio and visual perceptual parameters; one-on-one mapping on the technical level; or combination of the two. The actual "dance moves" are in my case always based on physics or pseudo-physics-based algorithms.

4.4 Texture-based Audiovisual Clusters

I use the term texture-based audiovisual clusters to refer to audio visual textures that are glued together with a high degree of cross-modal correspondences and intramodal perceptual grouping. Intramodal perceptual grouping, which is essentially an illusion that makes one connect concurrent elements of sound and image that are not structurally linked, can be amplified by adding actual synchronisation of certain audio and visual events. An example would be *Continuum* by Paul Jebanasam and Tarik Barri (Jebanasam & Barri, 2016). A similar example, although without any temporal synchronisation but with very explicit cross-modal correspondences, would be an excerpt from *Aurora* (1:11 to 1:38) by Ben Frost and MFO (Frost & Weber, 2014). In this excerpt we hear a *Shepard tone* cross-modally corresponding to constantly moving reflections of light. Such influences can be seen in my solo work *Capillaries Capillaries* (7:13 – 7:51) and in a collaborative project *Synspecies* (see *Synspecies - Spaceless Latitudes* between 0:32 – 1:00 or 5:25 – 6:14).

Ryoichi Kurokawa, on the other hand, uses audio noise to complement distortions of geometric objects – which changes the texture of the visual object. An example can be found in an excerpt from *ground.alt* (00:13 to 00:33) where various densities of visual noisiness are combined with different frequency bands of audio noise (Kurokawa, 2014). Similarly, in *Capillaries Capillaries,* I used rough 'electromagnetic' sounds as well as various sources of audio noise to complement the shaking of geometry and distorting geometry textures (12:45 – 12:58).

4.5 Switching Between Micro-isomorphic and Macro-isomorphic Approaches

Examples that exploit a wider range of the isomorphic spectrum are rare, for instance composed works that consist of very explicit micro-isomorphism (tight audiovisual objects), as well as very explicit macro-isomorphism (audiovisual clusters). One example would be *Seismik* from Herman Kolgen. *Perfo-extracts* video from *L.E.V. festival* shows a clean example of micro-isomorphic object between 1:15 to 1:53 (Kolgen, 2014). Putting the background ambience aside, the micro-isomorphism is formed from an extremely simple and reduced digital representation of a seismograph that corresponds to undulations and distortions of acoustic energy. The seismograph activity is interrupted with the moment of *synchresis*, caused by the change of visual scene at the moment of a transient based audio cue. The scene that follows is a macro-isomorphic world composed from an eerie drone and dark-ish landscape. Beside a strong cognitive complementation between the sound and the image, ranging from spiky resonant sounds and spiky visual shadows to similar movement between shadows and partials, camera follows the "swooshy" sounds with jolty moves which forms another layer of macro-isomorphism.

An example of switching between the two types of isomorphism in my work can be found at various times in composition *Singing Sand.* The transitions are often marked with *synchresis* as described in the *Seismik* example. Another method of moving between micro-isomorphism and macro-isomorphism is to morph between the two typologies. This is discussed in the following section.

4.6 Morphing Between Micro-isomorphic and Macro-isomorphic Approaches

Morphing between micro-isomorphism and macro-isomorphism is a constitutive element of the audiovisual language employed by Ryoichi Kurokawa. As far as I am aware, he is the only audiovisual artist that extensively exploits this compositional idea. In his later works he often uses macro-isomorphism to create a sort of suspense, where the time slows down almost to the point of standstill. An example can be seen in an excerpt (1:22 - 2:22) from audiovisual installation Unfold (Kurokawa, 2016). The excerpt starts with sustained tones accompanyed by slowly moving particles. The audiovisual material then exponentially accelerates into a highly saturated salvo of *synchresis* consisting of pulsating sounds and flashy visual elements. As we progress through the exerpt he continues wth simmilar type of gestures but starts adding micro-isomorphic elements. One could say that his gestures consist of exponential transitions between the stretched time of macro-isomorphism and the highly compressed time of micro-isomorphism and *synchresis* on the other. His micro-isomorphism is hence very short lived and also fragmented.

This influence is manifested in my works in various gestures where I quickly move between the two isomorphism. Examples in my works can be seen for instance in *Singing Sand* (4:13 – 5:00). My variation on the Kurokawa's influence generally produces less fragmented result. I tend to create slower and less intense gestures whilst putting more emphasis on the micro-isomorphic gesturality.

5 Composition Within the Audiovisual Isomorphic Paradigm

The aesthetics of my work are conditioned by the techniques I use. Before I begin discussing the elements of my compositions, I will briefly introduce the two paradigms that encompass the contrasting techniques I employ.

These paradigms are:

- The generative paradigm: using custom made tools with various levels of indeterminacy to generate content. Indeterminacy in most of these cases comes from the physics-based algorithms through which I simulate organic qualities of movement and then translate this movement into sound (see section 5.1.1), visual effects (see section 5.2.2) and temporal effects (see sections 5.3.1 and 5.3.4).

- The DAW paradigm: using tracks, audio and midi clips, automation curves, etc., to carefully shape sound and image (in my audiovisual system the visuals are also controlled via *Live*). The DAW paradigm enables me to partially control generative content and to juxtapose it with composed elements.

5.1 Exploring Audiovisual Hierarchies

Audiovisual hierarchies refer to the audiovisual relationships in terms of which medium leads and which medium follows during the process of composing. According to my experimentation with hierarchies, different hierarchies have different aesthetic implications. Creating gestures through the sonification of geometric motion, for example, generates results that I would not be able to conceive or even replicate if starting with the audio. Hence, it is important for me to explore the aesthetic potential of different hierarchies and, especially, to juxtapose the resulting audiovisual material inside single compositions.

Below, I present the three core audiovisual hierarchies, namely, *geometry leads audio follows*, *audio leads geometry follows* and *constant switching between audio and visual leads*.

5.1.1 Geometry Leads Audio Follows

When geometry leads and audio follows, sound or rhythm is derived from the real-time sonification of the visuals. The first stage of the workflow is to create an interactive visual object that can be later sonified and played like an instrument. The element of the visuals that I sonify are the moving vertices – I map the properties of moving vertices to audio parameters. Various aspects of such audiovisual objects are discussed in other parts of this chapter: in section 5.2.2 the audiovisual tightness is considered, while sections 5.3.1, 5.3.4 and 5.3.6 deal with temporal implications of the audiovisual hierarchy in question. Here, I will discuss the general aims and issues behind these audiovisual experiments.

The "geometry leads and audio follows" approach was predominantly used for the exploration of the sonic and musical potential of visual structures and their physicality. Movements of individual vertices were exploited for the implementation of generative rhythms (see *Capillaries Capillaries* 5:07 – 8:00), particle-based sound granulation (see *Singing Sand* 0:00 – 2:30) or geometric-undulation-driven additive, subtractive and wavetable synthesis (see *Capillaries Capillaries* 20:15 – 22:15). These approaches address questions such as: what sounds and internal rhythms can be extracted from geometric gestures or non-teleological undulations, how to sonify large amount of visual data without losing tight gestalt qualities, and how to intuitively control a large amount of sonic parameters via 3D animation. In order to address the question of how to deform the musical time with geometry, a special sequencer that treats the static temporal grid as a deformable elastic string was implemented (see section 5.3.1).

My global aim was to create unconventional audiovisual objects that would have some implicit aesthetic value *per se* and to compose with them – meaning that the overall aesthetics of a piece could emanate from these objects. I tried composing with the objects I implemented, using exclusively objects where geometry leads and audio follows, but I found the compositional possibilities too restrictive. And while I can appreciate the more alienated aesthetics that emanate from such constrained conditions in the works of others, such as the audiovisual works of Theo Burt or the piece *Trails* by Philippe Kocher and Daniel Bisig, I decided I wanted to have more expressive freedom in my own practice. Hence, I decided to try to integrate my geometry-driven audiovisual objects with influences ranging from techno music to acousmatic music. In order to do that, I had to adapt both sides – my geometry-driven objects as well as the genre-based idioms. An example of such mutual adaptation can be heard and seen in *Singing Sand,* in the excerpt between 5:21 and 6:16, where geometry-driven gestures are shaped to fit onto the grid of metre and harmony whilst a pre-composed "dub-tech" rhythmic pattern is distorted in the temporal sense according to the accents of the gestures. The result is a musical groove that emanates from the visual geometric gestures (see section 5.3.4 for more details).

The importance of audiovisual tightness in my works is discussed in section 5.2. Here, I will discuss a special issue regarding audiovisual tightness that relates to the sonification of larger number of particles or vertices. This issue was explored in the particle-based audio granulator (see the video *Particle-based Granular Synthesis - Test 2*) that became the core element in the piece *Singing Sand.*

When sonifying the particles in *Singing Sand* using a custom-made audio granulator, the issue was how to achieve audiovisual tightness despite the relatively large amount of visual information represented by 250,000 visual particles. The implementation of the granular sampler I used for sonification of the particles allowed me to map only up to 1024 particles to audio grains. After extensive experimentation, I concluded that sonifying only 128 particles gave me the best results. Sonifying more particles actually resulted in a lower gestural definition of sound and, hence, a lower degree of audiovisual tightness – presumably due to the phase cancellations, transient smearing and perceptual saturation of audio information. Hence, the problem was not in the technical limitation of my granular sampler as sonifying more visual information was not the solution. The problem was how to sonify only what is perceptually meaningful. I needed to somehow separate the visual particles that catch our attention from the rest, and sonify only them. Intuitively, I thought that I would need to sonify only the fastest particles, which can be a good solution for a more controlled motion, but in *Singing Sand* the motion is often chaotic and the fastest particles travel with an extremely high speed somewhere far away from the visual field of the screen. The final implementation of sonification thus aimed to sonify only the explicit visual gesturality in order to achieve tight micro-isomorphic qualities. That was achieved by employing a number of parameters that would determine the method of mapping the particle velocity to grain parameters. The parameters had to be set separately for different behaviours of visual material as I was not able to find a method that would work for all scenarios[8].

Sonification of geometric motion can, on the one hand, provide very tight micro-isomorphic relationships, while on the other hand, it can generate very tense audio

[8] After downsampling the original position matrix and calculating the velocities, the velocities were sorted from the lowest to the highest and split into five groups, each comprising of 128 velocities. Sometimes I mapped velocities to grain parameters from only one group, while often I used the difference between two groups of velocities in order to focus on the differences in motion. The latter method resulted in more "snappy" audio responses. The "snappiness" was additionally adjusted by fine-tuning the mapping curves of either logarithmic or exponential mappings.

material. The additional musical tension comes from audio phrases, rhythms and grooves that diverge from established musical idioms to which we are accustomed to listening. Hence, in order to create the most tense audiovisual material that I use in my compositions, I use the combination of the two elements.

5.1.2 Audio Leads Geometry Follows

The "audio leads and geometry follows" approach was used for composing visuals on top of pre-composed music or individual musical phrases. The aim here was to use the perceptual organisation of "visuals" to distort the perceptual organisation of sound and its meaning within pre-composed music.

My implementation of this approach can be split into two categories:

- Matching audio gestures with visual gestures in a predominantly micro-isomorphic manner. The workflow encompasses the preparation of visual material in advance and recording it as geometric *x-y-z* data (see section 6.3.3 for details). Recorded visual gestures are then matched with audio gestures through the manual manipulation of visuals using automation curves in *Live*. Additionally, the visual gestures can be modified by mapping audio analysis parameters to geometric *x-y-z* data which also emphasises micro-isomorphic qualities. An example can be seen in the *Synspecies – Teaser* video (between <u>00:15</u> and 00:37) where visual gestures were matched to pre-composed audio gestures while the amplitude of audio noise was mapped to the vertical stretching of the visuals in order to achieve tight audiovisual integration.

- Composing visuals that correspond to macro qualities of the music. This encompasses either synthesizing visual material directly for the music in question (see *Capillaries Capillaries* <u>16:18</u> – 17:33) or, alternatively, finding the best match from the pre-recorded geometric samples[9] (see *Singing Sand* <u>9:52</u> – 11:53). In both examples there was no editing done in post-production. What appears as editing is a combination of numerous visual parameters being changed at the same time which is all controlled from the timeline in *Live*.

[9] The term geometric sample refers to visual gesture.

The main difference between gestalts where audio leads and geometry follows, as opposed to gestalts where geometry leads and audio follows, is that the first approach enables the music to be composed without any visual constraints. That inevitably leads to different aesthetic outcomes.

5.1.3 Constant Switching Between Audio and Visual Leads

By constant switching between audio and visual leads, I am referring to workflows where audiovisual objects emerge from constant switching between sound design (while watching visuals) and visual design (while listening to the sound) – this approach thus demands visuals that run in real-time. The audiovisual object thus emerges from micro audio elements and micro visual elements being assembled across both modalities. The audio element can be either a type of an oscillator, filter setting or additional audio layer etc., whilst the visual element might be the amount of geometric distortion, brightness of visuals or type of a drawing primitive, etc. This approach does not prioritise any of the two mediums, as neither the sound nor the visuals are prepared in advance. However, each process needs to start somewhere, and the first building block can be either a sonic or visual element.

The aesthetic implications of constantly switching between audio and visual leads can be observed in all micro-isomorphic gestures in the section from *Capillaries Capillaries* (10:52 – 11:40), which are very tight. Also, the musical language in this section is very fragmented, consisting of short and contrasting gestures. Such fragmented musical language is in my case typical for this approach.

5.1.4 Combinations of Hierarchies

There are plenty of examples in my works where I combine either two or even all three of the hierarchies presented in this sub-chapter. Here, in this section, I will present an example where the cross-fertilisation of all three approaches is clearly evident.

I started composing the third section of *Singing Sand* (2:29 – 4:05) in a very classical EDM manner – with a kick drum. After I made the basic rhythmic pattern, I added the visuals. Since the visuals are sonified by mapping the velocity of visual particles to audio grains, the integration of the visuals needed to consider their musical part. The "original" sound of the visuals, with only subtle audio effects applied, can be heard towards the end of the excerpt (3:49 – 4:05) when this noisy layer gets gradually louder (the "original" sound is muted in the first part of the section). The sonic qualities and the groove of the sonified visuals, however, is also present in the first part of the section. There, the sound of the visuals has been subtly vocoded with the bassy pad sound while its groove was tamed with side-chain compression. The groove is also evident in the percussive elements (reminiscent of the sound of walnut shells) that also originate in the sonification of particles – another audio layer was recoded using different settings on the granulator resulting in percussive rather than continuous sounds. Hence, the musical backbone, excluding the kick pattern, was done in the manner of geometry leads and audio follows. The resonating overtones, emerging from the audio of sonified particles being vocoded with synth chords, determined the rest of the orchestration – additional melodies and chords were mixed in very quietly in order to interfere with overtones and thus function as a sort of a sonic hue that is slowly changing in the background. In response to the addition of harmony, I mapped the velocity of particles to their colour in order to achieve a better cognitive complementation. When searching for the appropriate spectrum of colours, I constantly switched between trying out different musical parameters (different chords, different settings on vocoder, reverb, etc.) and trying out different visual parameters (colours, combinations of different visual layers, different drawing primitives, etc.).

5.2 Audiovisual Tightness

5.2.1 Audiovisual Reduction

Tight audiovisual objects and loose audiovisual clusters define the spectrum of my (isomorphic) audiovisual relationships. Since I like to compose with strong contrasts, I need to work inside an aesthetic framework that allows me to achieve the desired audiovisual tightness. In this section, I will discuss the impact of audiovisual tightness on my overall aesthetics.

The most important factor when trying to achieve tightness with complex audiovisual material, at least according to my aesthetic preferences, is the reduction of audio and visual information. I decided to reduce visual information by rendering 3D geometry as a combination of points and lines in the negative space. Such visual objects can be reduced to a single point or a line if necessary, whilst the complexity can be increased by rendering the geometry with numerous different layers. Sounds that, according to my preferences, cognitively complement such visual material are various synthetic noises, "electromagnetic glitches", or what I perceive as cold digital clicks, etc.

This described audiovisual world represents my aesthetic starting point when composing, and with each piece in the portfolio, I was trying to stretch the framework of my aesthetic.

5.2.2 The Embryonic Audiovisual Object

The tight audiovisual object, powered by micro-isomorphic relationships, forms the centre of my compositional universe. Not only the starting point of my compositional process for each piece, it also represents the reference point for all the further variations of the audiovisual language inside a single piece. Thus, I think of such an audiovisual object as an embryonic audiovisual object.

In *Capillaries Capillaries*, the embryonic object can be seen for approximately the first five minutes of the piece. The audiovisual world consists of only a few points and lines which make sounds as they move and connect. These "clicky" sounds consist of different audio layers and the one that stands out the most is the sound of a fluorescent tube bulb[10]. Additional "orchestration" was made with the use of "electromagnetic sounds"[11]. I found these sounds appropriate to complement the visuals, which I find cold, harsh and "micro-electronic". An aspect of the visuals is also a set of flickering lights that correspond with the electromagnetic sounds. All these audio and visual qualities add to the cognitive complementation making the audiovisual objects tighter, as opposed to when using only spatio-temporal congruence. There are also tense jerky rhythms inherent to the movement of this embryonic object. These qualities are also part of aesthetic starting point of the whole piece and are reflected across the composition. At the same time, the audiovisual tightness of this object defines one extreme of the isomorphic spectrum used in the piece.

In the case of *Singing Sand*, I decided to use more complex audiovisual material as an embryonic object. In order to achieve the desired tightness, I had to also increase the complexity of the structural relationships between the sound and the image. I mapped the speed of individual particles to their colour and also to sound so that the speed of

[10] Fluorescent tube bulbs often make very short transient sounds from the moment they are turned on to the moment they warm up.

[11] Electromagnetic sounds are sounds that originate from the fluctuation of an electromagnetic field and were recorded with a DIY microphone consisting of a coil soldered to a mini jack. The source of the electromagnetic field was my mobile phone.

movement is reflected in sound as well as in colour. The result was an audiovisual object that possesses significantly more audiovisual information than the one in *Capillaries Capillaries* but maintains a relatively high degree of audiovisual tightness (see *Singing Sand* between 6:56 – 7:23). When designing the object, I used *Leap Motion* controller and pitch-bend wheel on a MIDI controller to interact with it. Using the pitch-bend wheel I was controlling the amount of particle attraction or repulsion while the position of my hand determined the position of the centre of attraction/repulsion. The axis of rotation and radial velocity of particles was controlled by orientation and a *grab strength*[12] of my palm. Initial experiments can be seen in video *Particle-based Granular Synthesis - Test 2.*

5.3 Time

The bulk of my experimentation with time revolves around the idea that geometry, conceived as points that move in 3D space, can be used to either generate or deform time.

Generating Time

The process of generating time is based on the sonification of geometric undulation which reflects organic qualities. Sonified undulation in my compositions can function as an unpredictable event generator – a sort of irregular metronome – in which the timing of events reflects the organic qualities of geometric undulation.

Deforming Time

In this approach, pre-composed musical phrases or sequences can be transferred onto a custom-made timeline which has elastic qualities and can hence deform these pre-composed temporal structures.

Other approaches deal with "sampling", time-stretching and time-freezing in the visual domain.

[12] *Grab strength* is a term used by *Leap Motion* developers to "indicate how close a hand is to being a fist" (Leap Motion, 2018).

5.3.1 Deforming Timelines: Stringvencer

The idea behind *Stringvencer* was to translate the organic qualities of physics-based geometric motion into deformations of time.

Stringvencer is a custom-made sequencer that enables deformations of pre-composed temporal structures. These temporal structures need to be based on MIDI notes and *Stringvencer* uses physics-based algorithms to displace these MIDI notes in time. More accurately, MIDI notes in *Stringvencer* are attached to a timeline which is deformable – the timeline behaves like an elastic string that can be stretched, tangled, thrown around or anything in between. This can be seen in the presentation video of the first version of the *Stringvencer 1.0* (Droljc, 2015).

Stringvencer works in combination with *Ableton Live*. Any MIDI sequence from *Live*[13], with a maximum length of 32 bars, can be imported into *Stringvencer*. Once the MIDI sequence is in the *Stringvencer*, it means that the MIDI notes are attached to a virtual elastic string. The amount of temporal offset caused by string deformations, can be controlled for each individual MIDI note. A newer version of *Stringvencer*, *Stringvencer* 2.0, also offers an additional temporal offset with random values. Similarly, the random offset of MIDI notes can be controlled separately for each MIDI note. If the original MIDI sequence is based on metric time, *Stringvencer* can be used to add swing by leaving certain MIDI notes at their original position whilst offsetting others. An example can be heard in the groove of the music used in the *LPM Promo Video* (see *LPM Promo Video*).

I initially implemented the *Stringvencer* tool because I wanted to be able to instantaneously deform time. The realisation of this idea can be seen and heard in the presentation video *Stringvencer 1.0*, where the string as a timeline is thrown around (Droljc, 2015). In such cases, the time is deformed to an extent that the sequence gets restructured. This idea was also tested in the musical piece *Fail Better*, where I recorded the temporal deformations back into *Live* as MIDI notes. The resulting gestures were edited and processed in *Live*. The aim of processing was to enhance the effect of temporal deformations by making the amount of audio distortion and pitch modulation of a synth roughly correspond to the temporal deformation (*Fail Better* 8:05 – 9:28). There is, however, a problem with such ways of using the tool – when the string is thrown around or tangled, we cannot perceive the deformation of time. What could be considered as temporal deformation on a technical level appears as a random musical gesture. One way

[13] MIDI sequence can be also created by converting an audio sequence into a MIDI sequence. Alternativelly, any MIDI sequence can be imported into *Live*.

to solve that problem would be to compose a piece, in which the visual element of the *Stringvencer* would be integrated into the visuals of the piece. However, I decided to avoid that option and to rather explore different ways of using the tool – to either create grooves or to deform the temporal structures very slowly.

In *Capillaries Capillaries*, between 17:15 – 19:40, there is an example of the gradual temporal deformation. A temporal structure based on 4/4 rhythm, established in the section before the beginning of this excerpt, starts to gradually fall apart. The backbone of the temporal deformation was created with *Stringvencer* by exciting the string very subtly in various parts using the mouse. This process was improvised and the output from *Stringvencer* was recorded in real-time as MIDI data in *Live*. The sequence consisted of eight rhythmic layers that were recorded on eight MIDI tracks in *Live* (see Figure 2). The recorded MIDI sequences were then additionally edited in *Live* by filling certain silences with concatenated MIDI notes (short MIDI notes to create a "salvo" of sounds) or by prolonging the silences where additional suspense was needed. Part of editing was also to occasionally mute layers in order to avoid the perceived sonic saturation. The only layer that got significantly edited were the hi-hats where I inserted a number of equispaced notes in between either two existing hi-hats, between a hi-hat and a snare or between a snare and a kick. Such hi-hat pulsations, although mixed very quietly, provided a stable rhythmic reference against which the temporal deformations were made clearer. That helped to increase the musical tension where needed.

The section between 11:41 and 14:45 in *Capillaries Capillaries* was done in a very different way. Here, I was looking for a type of temporal distortion that would preserve the groove, even when the time would be significantly distorted. After a number of experiments, I realised that one way to achieve that would be to constantly move between more and less distorted rhythms while gradually increasing the ratio in favour of the more distorted rhythms. Therefore, I created four different temporal deformations, ranging from mildly to heavily distorted. Each of them had a different tempo, as heavily distorted rhythms appear to us, or at least to me, much slower than they really are. The rhythm with the lowest amount of distortion was 65 bmp while the rhythm with the highest amount of distortion was slightly more than twice as fast at 132.1 bpm. I implemented a function to morph between these four presets and mapped the morph parameter to a slider on a MIDI controller. The slowly changing groove in the section between 11:41 and 14:45, came from switching and cross-fading between the four presets. In order to do that, I had to modify the *Stringvencer* (this addition will be available in a separate application that will accompany the upcoming version of *Stringvencer* (3.0)). The morphing between the presets was recorded in real-time as automation in *Live* and additionally edited (see Figure

3). Syncopations were created by switching to syncopated variations of the original MIDI sequence inside *Stringvencer*, as the *Stringvencer* 2.0. allows the importing of multiple sequences for each MIDI channel.

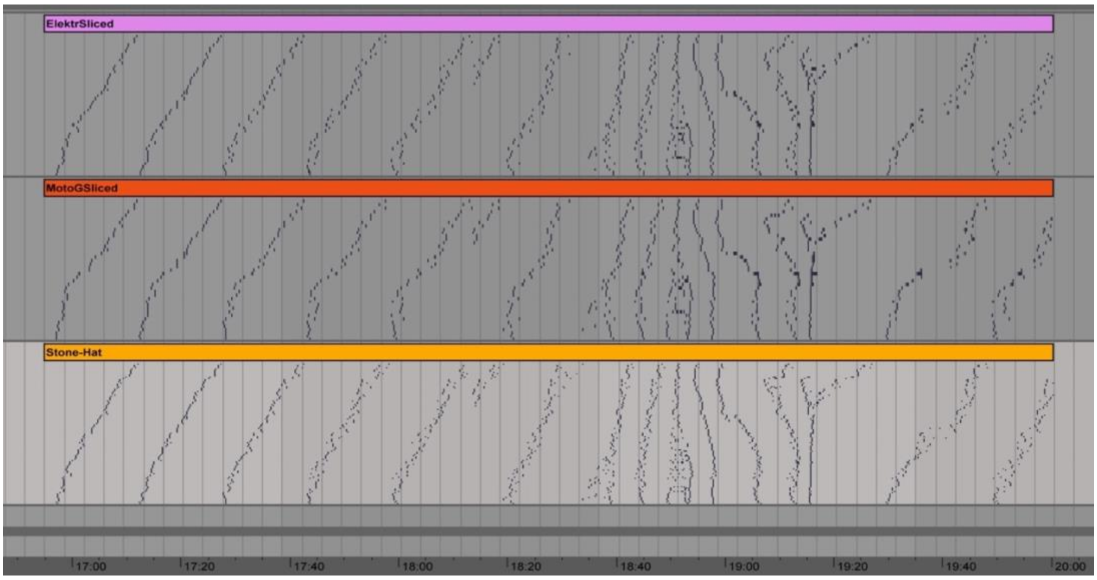

Figure 2: The gradual temporal deformation of rhythmic patterns in *Capillaries Capillaries* between 17:00 and 20:00. Temporal deformation was achieved by employing the *Stringvencer* tool. The figure shows the recording of three MIDI tracks recorded directly from *Stringvencer* into *Live*.

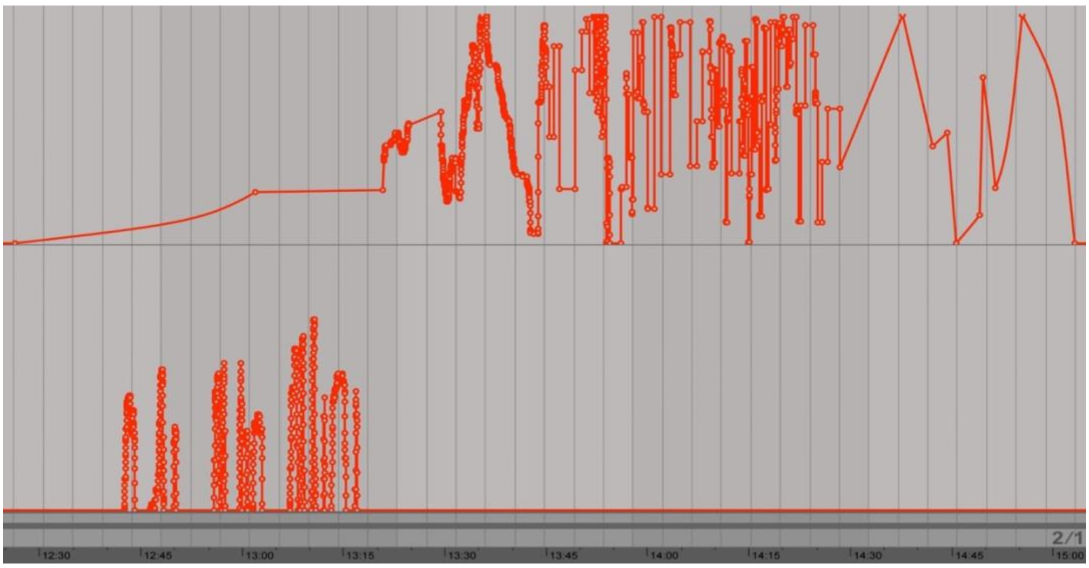

Figure 3: Automation curves in *Live* controlling the cross-fade between the four rhythmic presets (each preset having different degree of temporal deformation and different tempo). These two automation curves were added together in *Max*, hence they both target one cross-fade parameter.

5.3.2 Continuous Time vs. Fragmented Time

Capillaries Capillaries and *Singing Sand* both explore the juxtapositions of continuous and fragmented time. However, in *Capillaries Capillaries* the emphasis is on the exploration of continuous time while in *Singing Sand* the emphasis is on the exploration of fragmented time.

Continuous Time in Capillaries Capillaries

One of the aims behind *Capillaries Capillaries* was to compose a piece that would reflect the qualities of organic growth and decay. Hence, I wanted that certain elements of audiovisual material would in certain sections continuously morph between different states. Also, my preference was to implement continuous transitions between the sections.

An example of continuous transformation in *Capillaries Capillaries* is the excerpt between 11:41 and 14:45, where an audiovisual gestalt slowly decays over time and towards the end of the excerpt morphs back into a tighter audiovisual entity (the transformation includes various states of audiovisual material such as visual "dancing" object, audiovisual object and audiovisual cluster). This journey is interrupted with radical momentary deformations that are elastic (reversible) in nature. In order to focus on continuous deformations, I decided to use a very long time-window. Throughout the whole excerpt there are no edits and no step transformations of the visual scene. All momentarily transformations of the audiovisual gestalt hence needed to appear as deformations rather than spatio-temporal discontinuities. In a non-referential audiovisual world, spatio-temporal discontinuities can be introduced even by switching between different cameras, as a complex visual object might from different angle appear as different object.

Fragmented Time in Singing Sand

In one of the early stages of composing *Singing Sand*, before I started working with a timeline, I created a number of continuous and unrelated transformations of audiovisual gestalts. The idea was that these continuous transformations would represent parallel narratives, but the trajectory of the piece would jump between different fragments of these parallel stories. The only common denominator of these unrelated narratives/transformations are the similarities between audiovisual articulations that come from using a single custom-made visual synthesizer.

Although this initial idea has a clear impact on the way that time is treated in the composition, the underpinning idea is not very evident when experiencing the piece. I believe there are two reasons for that. One reason is the level of abstractness in the source material and the degree of transformations (some of the source material can be observed in any of the longer time windows in *Singing Sand*, such as the excerpt from 0:53 – 2:30). In order to make the initial idea more apparent, I would need to either use stronger and more memorable gestalts or treat the existing material more carefully. The reason why I did not explore the latter option is that I decided to sonify the visual particles and use them like a musical instrument. Therefore, I needed to concatenate temporal fragments of visuals in a way that made musical sense – sequences of visual time-windows had to create musical phrases that could be integrated into the larger musical whole and during the compositional process this became my main priority. Hence, the final visual fragmentation, which I understand as temporal fragmentation, is predominantly subordinated to the musical needs. However, such musical gestures also needed to look good, so there was always a compromise between what looks good and what sounds good in a given context. During the compositional process there were numerous occasions where the visual gesture sounded exactly as I wanted it but it was unacceptable for visual reasons (and vice versa).

5.3.3 The Audio-visual-time Object

In section 5.1.1., I focused mainly on the musical properties of temporal distortion. In this section, I will present an idea that inspired *Stringvencer*: the idea of an audio-visual-time object. In an audio-visual-time object, the sound, image and time all correspond.

When thinking about audio-visual-temporal transformations, I came across "Three Studies for Figures at the Base of a Crucifixion" by Francis Bacon. The triptych represents mutilations of the human form. What caught my eye was the fact that these figures were still recognisable as distortions of the human form, despite being severely distorted. This made me think about the importance of using strong gestalts, such as human bodies, when working with severe distortions. I wanted to translate this idea from visual art into an audio-visual-temporal domain, in order to create a strong audio-visual-temporal gestalt that could be severely distorted without having to lose the recognisable qualities of the original state. From my experimentations with *Stringvencer*, I knew that complex rhythms and noisy or "glitch" sounds would not make good material for a strong gestalt, as the temporal distortion easily turns them into textures or random sounding rhythms divorced from their origin. Hence, I decided to use classic EDM sound elements such as kick and snare drum and placed them in very simple patterns. The accompanying visual object also had a clearly recognisable shape. The result, including the continuous deformation, can be heard in the previously discussed excerpt from *Capillaries Capillaries* (11:41 – 14:45).

The implementation of the audio-visual-time object in the excerpt in question does not encompass a radical sonic deformation. Here, I did not want to detune the pitch of the pitch-based sounds, which is what I did in *Fail Better*. Therefore, I decided to portray the temporal deformations only on the visual object, while using sound as the source of visual deformations. Accordingly, as the temporal deformation becomes more severe, the sound causes more deformation to the object – elastic deformations become plastic (irreversible) and the frequency of the sounds causing the deformation is increased. The majority of the plastic deformations comes from the pulses of the snare drum that were added as part of the editing process of a MIDI recording[14]. Besides the obvious micro-isomorphic correspondences between sonic energy and visual deformation, the musical "orchestration" also follows and supports the drama. As the visual object becomes noisier and more dispersed, so does the music.

5.3.4 Getting the Groove from Visual Geometric Gestures

The protagonist of the piece *Singing Sand* are sonified visual particles[15]. These particles represent the only visual element appearing in the piece, hence the visual gestures are rendered as micro-isomorphic audiovisual gestures[16]. Throughout the beat-based sections of the piece, I was trying to fit these musical gestures emerging from visual particles into a metric structure by time-stretching and time-compressing the visual material (see section 5.3.5 for more details on visual time-stretching). My aim was to keep the organic qualities of the visuals in the sound, therefore, I made the micro-isomorphic audiovisual gestures only partially fitting the metric grid. To fix the imperfections of temporal synchronisation I decided to deform the metric grid – onsets of sounds representing the "orchestration" were positioned according to the accents of the micro-isomorphic audiovisual gestures. This created the musical groove that was rooted in the visual gestures[17].

[14] See section 5.1.1 for the recording process.
[15] Individual visual particles are mapped to individual audio grains of a granular sampler. For more details please refer to section 5.1.1.
[16] Musical gestures emerging from sonification of visual particles in *Singing Sand* are at times muted.
[17] In the beat-based parts of *Singing Sand*, musical gestures emerging from the visuals are at times exposed (mixed loudly), at times blended with the overall sound world (mixed quietly), and at times muted. In all three cases these musical gestures were used to define the groove of the 4/4 rhythm used in the piece.

One implementation of this approach is to temporally deform the pre-composed rhythmic pattern according to the musical accents of the micro-isomorphic gesture. An example would be offsetting the kick and the snare drum from the temporal grid as it can be seen in *Singing Sand* between 5:59 and 6:39*.* The musical accents of the micro-isomorphic gestures are sometimes synchronised with the drum sound (6:11) while sometimes the drum sound would be placed against it (6:02). These musical accents often fall on the border between two time-windows, which looks like a point of edit, although technically speaking, there was no editing done in *Singing Sand*[18]. When creating the groove, I was also trying to match the energy of the visual gestures with the energy of music, thus the resulting groove represents a compromise between micro and macro audiovisual parameters.

To spread the micro-rhythm of musical gestures emerging from the visuals across various audio layers of the overall "orchestration", I vocoded these layers with the source sound. This effect of vocoding can be heard throughout the majority of the piece (in all the sections with coloured visuals). The effect is most noticeable in parts that lack pulse and therefore also groove. In the parts with beats and metric time, the internal rhythm of these musical gestures adds to the overall groove of the micro-rhythm. The micro-rhythm of audio gestures perfectly complements the distorted rhythmic patterns described in the previous paragraph as the distortion of rhythmic patterns was derived from the same audio gestures. An example can be observed between 5:20 and 6:06. In this example, as well as throughout most of the piece, each audio layer had a vocoder with different attack and release settings. On the bass track, for instance, I used slow attack and long release settings which resulted in a slew limited spectral following which only vaguely reflects the visual action. Additionally, I drastically reduced the resonant frequencies when mixing so that the effect is very subtle. On sounds with an emphasis in the mid and high-mid frequency range, I used shorter attack and release settings in order to also emphasise the micro-isomorphism between sound and image.

[18] What appears as editing in *Singing Sand* is a combination of numerous visual parameters being changed at the same time which is all controlled from the timeline in *Live*.

5.3.5 Extreme Visual Time-stretching

In order to have more accurate control over the articulation of visual geometric gestures, I implemented a time-stretching algorithm for manipulating recorded visual gestures.

When a video file is played back at a slower speed than at which it was recorded, the result can be "choppy" playback. While contemporary software for video editing offers various solutions, such as optical flow or frame blending, they all have their limitations. Since I very rarely compose in video editing software, I needed to find a solution for time-stretching my visual gestures in *Max*. The solution needed to be real-time and CPU friendly, as all the pieces in the portfolio were composed on a Mac laptop from 2013.

The solution I chose is based on recording frames of x-y-z geometry data and playing them back with a frame-interpolation technique. Accordingly, each recorded frame does not contain the RGB data needed for the reconstruction of a digital image, but rather a position matrix consisting of x-y-z position data for each individual vertex. Using a frame-interpolation playback technique enables me to do extreme time-stretching of the recoded geometric gestures in real-time.

This technique has a significant impact on my visual and audiovisual aesthetics as the recorded gestures can be played back or reversed at any speed. The speed of playback is controlled with automation curves in *Live* – the steepness of the automation curve determines the speed of the playback while the descending curves result in reversed playback (automation value 0 refers to the first frame in the sequence while value 127 refers to the last frame). Therefore, the steepness of the automation line that would result in original playback depends on the number of frames in a recorded sequence. When composing, I did not know which steepness would give me the original playback speed; that was my conscious decision when implementing the tool. The reasoning behind this decision is that this limitation of the tool forces me to not prioritise any specific playback speed and, therefore, instead, to always seek the playback speed that best matches the energy and articulation of the music (or resulting sound, in the case of geometric sonification). Additionally, since I am recording only position data, the played-back gestures could be scaled, rotated, translated, coloured or rendered in any desired way.

This technique for matching visual gestures with music was used between 3:19 and 4:19 in *Synspecies – Spaceless Latitudes*. The music in this part is very tense and articulated and I wanted to amplify the tension with predominantly micro-isomorphic audiovisual relationships. Having a careful time control over the playback, and hence the articulation of recorded visual gestures, enabled me to achieve the audiovisual tightness I wanted.

5.3.6 Shaping Non-teleological Rhythms

The rhythm in the first movement of *Capillaries Capillaries* (0:00 – 9:00) comes from the sonification of a visual algorithm that connects neighbouring points with lines when they get close enough together. Each connection triggers a sound and, similarly, each disconnection, when a line gets erased because of excessive distance between the points, triggers a different type of sound. An example of the rhythms that come from such sonification can be seen in the video *Rhythm of Undulating Geometry*. In this example, the visual algorithm was applied to only 9 points (vertices) of an undulating 3D object. For more information on technical details, see section 7.2.1.

The rhythm in the example video *Rhythm of Undulating Geometry* is non-teleological. By non-teleological, I mean that it does not have any specific direction – the rarefactions and compressions, which can be understood as tension and release, do not have a temporal structure which makes the rhythm point in various directions. At the same time, the rhythm reflects some organic qualities that I believe come from the visual undulation. I decided to use this rhythm in the piece *Capillaries Capillaries*. I recorded the rhythm as MIDI data, which I then used to trigger kick drums in the section between 5:07 and 6:37 as well as lights and accompanying "electromagnetic" sounds in most of the first movement (3:42 – 9:00).

The challenge in this case was how to build an abstract narration on top of non-teleological rhythm with as little editing as possible[19]. The main elements I used to create a narrative were superimposed audio layers with clear direction and similarly directed automation of the parameters that controlled the audio processing of the rhythm – an example would be the filter sweep on the kick track at 5:43.

In the first part of the first movement of *Capillaries Capillaries* (0:00 – 5:07) I tried to shape the non-teleological rhythm into phrases by controlling rarefactions and compressions. At the beginning of the piece (0:00 – 0:55), the rhythmic phrasing comes from controlling only one parameter – the threshold distance between neighbouring points

[19] Editing was mostly done by erasing unwanted notes while in the section 5:25 – 5:47, I added another kick drum following the "pulse" I perceived in the original non-teleological rhythm. The additional kick drum was recorded by me manually by triggering the sound on every perceived "beat".

which determines if the points will get connected or not. When the distance between two points is less than the threshold specified, the two points connect. When the threshold distance is set to zero, no connections can be made as points never get next to each other. On the other hand, increasing the threshold distance increases the chance of forming a connection. Adjusting the threshold distance thus allowed me to have total control over silences and partial control over the density of the events when connections started to form.

In the following part (2:40 – 3:40), the distance parameter was fixed while I manually controlled the speed of the flow and the scale of procedural noise, which is responsible for the underlying undulation. When the speed of the flow equals zero the undulation stops which results in silence. The scale of procedural noise on the other hand affects the distance between the points.

By varying all three parameters from both of the previous examples, I created rhythmic audiovisual phrases that can be heard and seen between 3:40 and 5:07 in *Capillaries Capillaries*. Around 4:42, the undulating shape was additionally distorted with "electric shocks"[20] which in return generated the rhythmic compression. The function of this rhythmic compression is to increase the tension before the scene drastically changes – black background switches to white.

5.3.7 Dynamic Audiovisual Relationships

In isomorphic paradigm, audiovisual objects cannot be cross-modally dismantled by incongruences of temporal structures between sound and image. However, changing audiovisual relationships across time can break the stasis of audiovisual parallelism. This concept I call dynamic audiovisual relationships and was presented in section 4.2. In this section I will analyse an example of dynamic audiovisual relationships from the piece *Capillaries Capillaries*. I will focus only on one audio element and its different visual correspondences across time.

In the excerpt between 2:40 – 3:42 in *Capillaries Capillaries* the high pitched sine wave sound forms three types of audiovisual relationships with the image: micro-isomorphic, macro-isomorphic and *synchresis,* where *synchresis* refers to the synchronisation of short

[20] Geometry was distorted by offseting the position of vertices with pseudo random values. The amount of offset corresponded with the volume of acompanying "electromagnetic" sound which appears to me like an audiovisual response to an electric shock.

bleeps with a sudden change in the image. In the excerpt in question we hear the sine wave sound nine times. As we move across the excerpt the sine wave:

1. marks the end of a visual transformation.
2. announces the next visual transformation.
3. marks the end of a visual transformation.
4. marks the step transformation and becomes micro-isomorphic with the visual object inside the cube.
5. marks the frozen time inside the cube and rotation of the cube.
6. marks the change in direction of particle movement.
7. no direct relation to visuals – purely a macro-isomorphic element.
8. no direct relation to visuals – purely a macro-isomorphic element.
9. back to its initial role – marks the visual transformation.

5.4 Compositional Indeterminacy

Indeterminacy in my works emanates from algorithms that can replace my compositional decisions or serve as an extension of my human capabilities and ideas. My aim is not to remove myself from the compositional process but, on the contrary, to put myself in a curatorial role, in which I can decide which resulting material can be used or not, what needs to be adjusted and what is ready to become a part of the piece exactly as it is. This enables me to experiment with various aesthetic deviations which I would not be able to implement on my own – either due to the excessive complexity of the processes involved or due to the limitations of my established thinking patterns in regard to composition.

There are three types of indeterminacy in my works – unrepeatable pseudo-randomness, repeatable pseudo-randomness and serendipity. The latter in my works refers to anything between "happy mistakes" and the results of physics-based algorithms used in unpredictable ways.

5.4.1 Algorithmic Drawing Based on Probabilities

In *Capillaries Capillaries*, between 6:17 and 8:42, there is a section that is reminiscent of a painting. It was made by the accumulation of video frames in which each frame was drawn on top of the previous ones. I see this process as accumulation of time in the present rather than moving on into the past. This extended present gives the listener-viewer a chance to explore the visual scene at their own pace.

Due to the accumulation of video frames, anything that moves on the screen leaves a trace behind. The only things that actually move in the excerpt in question are dots and lines. Since they are grouped together, their traces are reminiscent of brush strokes. Therefore, I will refer to this group of points and lines as the "brush". The movement that is leading the "brush" comes from the procedural noise which is a pseudo-random function with a repeatable outcome. I left the function freely running with the aim of making the "brush" strokes unrepeatable. Unpredictable micro gestures that originate in procedural noise determined the exact position and behaviour of the "brush". The parameters regarding the macro position of the "brush", such as x and y coordinates, were, on the other hand, controlled by me via automation curves in *Live*. In the same way, I controlled the speed and the type of the "brush", in which the type refers to the type of a *drawing primitive* used. The position of the lights on the other hand, which plays an important role in terms of what gets drawn or not, is entirely controlled by a pseudo-random function used in unrepeatable manner. However, the x-y-z positions of all the visual elements in the excerpt in question are limited to fit within the boundaries of a rectangular parallelepiped that frames the visual activity. In a similar way, the sounds that accompany the flickering of the lights are also unpredictable – these sounds represent randomly chosen audio fragments from a longer sample. The results have an unpredictable gestural and timbral variety.

A crucial element in getting a satisfactory result was controlling the x and y parameters of the "brush". When recording the scene into a video format, I made three different recordings and picked the one I preferred the most. All three versions of this section were recorded in real-time using an external video capturing device. A small part of the chosen version needed to be cut out, as lots of frames were dropped at once, for an unknown reason. In order to hide the technical failure, I integrated the glitch into the composition, by filling the gap with a still image. The result can be seen at 7:13. It adds a tense and unexpected twist to the otherwise continuous flow which functions well in the context of the piece. The integration of this technical glitch improved the excerpt in question.

5.4.2 Local Deformations of Visual Shape Controlled by Unrepeatable Pseudo-random Function

In *Capillaries Capillaries*, between 12:58 and 13:58, it appears as if the energy of the kick and the snare is causing plastic deformations of the visual object. On a technical level, each MIDI note causes an unrepeatable pseudo-random x-y-z displacement of a point that repulses particles. This repulsive power is strong, but very short lasting as it gets turned off quickly – similarly to a transient in the audio world. The destructive effect of particle repulsion varies from unnoticeable to significant as it depends on the distance of the point of repulsion from the object. This situation is reminiscent of a random "machine gun" shooting at the 3D object or perhaps some random destructive mechanical industrial process, which are the associations I wanted to evoke. Therefore, I left the sound of the pulsating snare, which is causing the majority of the deformation, very mechanical by mostly repeating one and the same sound at regular intervals.

5.4.3 Serendipity

Serendipity, as described in the introduction of this sub-chapter, is an indispensable element of my audiovisual work.

An example of serendipity would be an excerpt from *Singing Sand,* between 2:13 and 2:23, where the development of the visuals, which are also generating the sound, comes from an intersection of physic-based algorithms, logical operators and mathematical functions. The outcome was unintentional, and it represents a result of various factors that each shape the material in their own way.

An example of serendipity from *Capillaries Capillaries* can be seen in the excerpt between 17:33 and 17:53, in which the spatial position of a particle attractor, which also rotates particles, is being quickly modulated on a micro and macro level. This results in a serendipitous web of curves and spirals that were not intended when composing the piece. Serendipity is also echoed across the temporal domain as the distortion of visuals corresponds with the distortion of time. This distortion of musical time was similarly implemented using physics-based algorithms – by exciting a physics-based string that serves as a timeline (see section 5.3.1 for more details).

When generating diverse visual material without any specific aims, my approach to the synthesis of visual geometry is essentially the same as when using an audio synthesiser for experimental sound design. I tend to create a complex interdependence of parameters in which a small change of each parameter can cause unexpected changes, which are for me often impossible to understand or predict. The result, often very fragile, often emerges from the complex interdependence of parameters that needs to be set very precisely.

To me, serendipity represents one of the most important aspects of my work, as it enables me to explore the audiovisual worlds that lie beyond my imagination.

5.4.4 Loose Compositions

The use of the numerous techniques discussed in this sub-chapter makes my compositions partially indeterminate. Another layer of indeterminacy comes from the frame-rate fluctuation of real-time visuals in combination with physics-based algorithms. The frame rate in my works can fluctuate drastically. This is because when I compose, I mostly work with maximum CPU loads as I tend to use lots of CPU expensive processing. In *Capillaries Capillaries* the frame rate was fluctuating between approximately 10 fps and the desired 60 fps. Since physics-based algorithms are based on feedback-loops, their effect on the geometry depends on the number of iterations. In my works, in which geometric motion can be also used to generate sounds and time, the fluctuation of the frame rate also affects the audio and temporal domain.

The fluctuations of the frame rate are relatively consistent when playing back the composition in real-time. This results in only slightly different visuals which can be easily ignored when composing, but the differences are much more distinct or even disturbing in the audio and temporal domain. Hence, I need to adjust my aesthetics according to the available conditions. My compositions need to function even when the sounds and temporal events are not precisely fixed.

When rendering the visuals in offline mode with a fixed frame rate, the difference in the effect of physics-based algorithms can be extreme. Hence, for offline rendering, I need to apply dynamic correction factors to the parameters of a physics engine in order to compensate for the difference in frame rate. The result of the offline rendering hence always varies in comparison to the real-time renders. If the offline render appears unacceptably different, I adopt the correction factors until I am satisfied with the result.

This combination of technical obstacles and desired indeterminate elements has forced me to work with what I call loose compositions. Loose compositions become fixed only when recorded, whilst during the compositional process each playback reveals the inconsistency of audiovisual material.

5.5 Audiovisual Escapology

Audiovisual escapology, a concept I came up with in order to talk about my work, is a cross-modal extension to the musical idea of tension and release.

In general, *escapology* refers to the art of escaping from restraints or confined spaces. In this context, *audiovisual escapology* refers to the sound and image being tied as an object whilst trying to liberate themselves from the object. Constraints are hence represented by the micro-isomorphic audiovisual relationships which I use to amplify musical tension. The *audiovisual escapology* is about using the combination of micro-isomorphic and macro-isomorphic elements in order to create a feeling of entrapment as well as a tendency of both sound and image towards mutual liberation from the object. Cross-modal liberation is underpinned by macro-isomorphic relationships that provide a significant degree of temporal freedom (see section 3.3.4 on temporal freedom in regard to macro-isomorphism) whilst amplifying a sense of musical release.

The drama of *audiovisual escapology* revolves around entrapment (a tight audiovisual object), attempts at escaping (the interplay of objects and clusters), quasi-liberation (loosening the tightness of audiovisual objects or tightening of audiovisual clusters) and liberation (a loose audiovisual cluster). To simplify the discussion on *audiovisual escapology,* when referring to my work I will introduce two additional concepts, namely, *monoids* and *multiplexes.* A *monoid* is an extension of an audiovisual object and encompasses the quality of tension (entrapment) and the behaviour of the object (trying to escape). A *multiplex* is an extension of an audiovisual cluster and encompasses the quality of release (the liberation from the object-based entrapment into an abstract world that lacks concrete referents such as time and space).

5.5.1 Monoids

Monoids are like audiovisual "dough". These gestalts form clear shapes and can undergo elastic (reversible) and plastic (irreversible) deformations. As the "yield point" of the material is dynamic – *monoids* can hence endure "infinite" elastic (reversible) stretching or plastically (irreversibly) deform when subtle forces are applied. That gives them the ability to slowly evolve or disintegrate, while being temporarily interrupted by extreme forces (see *Capillaries Capillaries* between 11:38 and 12:37). These temporal deformations can be based on either micro-isomorphic or macro-isomorphic audiovisual relationships. Therefore, a *monoid* can be understood as an audiovisual object that can be temporarily transformed into an audiovisual cluster but then reassemble itself back into an object.

The structure of *monoids* spans from reduced one-to-one audiovisual mappings to complex audiovisual conglomerates where a multi-layered visual object cross-modally corresponds with compact mix of auditory streams – as long as the gestalt appears predominantly as an object. Hence, it is important that the visuals have the appearance of an object.

5.5.2 Multiplexes

Multiplexes consist of multiple visual objects grouped together in a parallel or serial manner and glued together with a continuous audio gesture or any other auditory stream. The continuity of the auditory stream gives the sense of connectedness, while the fragmented visual element contributes to the separation of sound and image from a single object and hence adds to the multiplicity of meanings, typical for our inner world of associations as opposed to object-based "semi-concrete" reality. *Multiplexes* could be described as audiovisual phrases where each visual element represents a different perspective or interpretation of some elusive multidimensional object or a feeling. An example would be an excerpt from *Singing Sand* (9:52 and 10:29) that functions in the piece as a "total liberation". Although *multiplexes* are predominantly macro-isomorphic, they can also encompass micro-isomorphic relationships. An example can be seen in *Singing Sand* (5:20 – 5:40) – this part functions as a "partial liberation" as micro-isomorphic relationships are clearly evident. Other examples include a visual diptych or triptych (see *Synspecies – Spaceless Latitudes* between 5:26 and 5:45).

5.5.3 Escapology in Practice

Capillaries Capillaries and *Singing Sand* can be described through the prism of *audiovisual escapology.* Both pieces start with a sense of entrapment (a tight *monoid*), reach a point of liberation (a loose *multiplex*) whilst the transitions between these two extremes in both pieces explore different trajectories. *Audiovisual escapology* can be also applied to individual sections of *Synspecies – Spaceless Latitudes* although this piece avoids the liberation stage.

5.5.3.1 Entrapment

The sense of entrapment represents one of several factors that contributes to audiovisual tension. Other factors were presented in sections 3.3.3 and 5.1.1. In section 3.3.3, "The Compositional Function of Micro-isomorphism", I discussed how audiovisual tension in my works predominantly relates to amplifying musical tension with micro-isomorphic relationships. At the end of section 5.1.1, "Geometry Leads Audio Follows", I discussed how the sonification of geometry in my works additionally increases the tension. For the implementation of *monoids* when composing, I use the combination of all these factors[21].

I try to achieve a sense of entrapment by explicitly exposing micro-isomorphic relationships and using occasional macro-isomorphic elements to create a contrast. Macro-isomorphic contrast in such cases functions primarily as a reference point for tightness. Such *monoids* can be observed at the beginning of works *Capillaries Capillaries* (0:00 – 1:00; 2:40 – 3:40) and *Singing Sand* (0:00 – 0:53). The main difference between the two examples is that excerpts from *Capillaries Capillaries* focuses on one *monoid* using longer shots while the excerpt from *Singing Sand* is based on the chain of briefly appearing *monoids* interrupted by macro-isomorphism. In the latter case, I was using short time-windows to create a perplexing scene by preventing the listener-viewer from fully grasping individual *monoids*. The perplexing effect in this case hence supports the tension.

The occasional macro-isomorphic elements mentioned in the previous paragraph in most of the cases represent cross-modal attempts at escaping. Entrapment is hence closely related to the attempts at escaping that will be presented separately in the next section.

[21] I use the sonification of geometry only in my solo works.

5.5.3.2 Attempts at Escaping

Attempts at escaping refers to explicit macro-isomorphic attempts at liberating sound and image from the object. These attempts are brief and in the long run always fail. Attempts at escaping can be understood as momentary transformations of audiovisual objects into clusters and back into objects. The transitions are often of an explosive nature and start with increased tension – as if a tension is needed for breaking the ties. This interplay of tension and release characterises the attempts at escaping which is a typical behaviour of *monoids*.

The section between 11:40 and 14:45 in *Capillaries Capillaries* begins with a *monoid* that has concurrent micro-isomorphic and macro-isomorphic qualities and is due to the latter slightly loose. Such audiovisual relationships are ready to be either tightened or loosened, which can be used to introduce some dynamic interplay of audiovisual tension and release.

In *Capillaries Capillaries*, from 12:15 – 12:35, there is a sequence of musical gestures, that demonstrate attempts at escaping. The first musical gesture consists of tense and abrasive electrical sound that emphasises the tension introduced by syncopation around 12:21. In order to additionally amplify the tension with the visuals, I "tightened" the *monoid* by creating a tight micro-isomorphism between the "electromagnetic" sound and the visual object – the visual object reacts to sound as if being exposed to an electric shock. The following gestures breaks the ties with the momentary macro-isomorphism (*Capillaries Capillaries* 12:22). We hear a "one shot" on a downbeat, with a long reverb tail applied to it, accompanied by visuals that instantaneously stretch the *monoid* into "infinity". This macro-isomorphic audiovisual gesture implodes back into an object and then the next attempt at escaping takes place. After numerous attempts the *monoid* becomes very loose and eventually disintegrates into macro-isomorphism (around 14:00).

A similar interplay of tension and release that we can observe on the micro time scale in *monoids* is also reflected on the form of the piece. The example from previous paragraph, for instance, continues with a "tightening" from macro-isomorphism slowly back into *monoid* (14:40). What follows is again a series of attempts at escaping that gradually loosens the *monoid* all the way to macro-isomorphism (around 16:10). This part functions as a "fake end" to the piece as the energy of the *monoid* completely dissolves. A moment later, we are thrown into a *multiplex* consisting of energetic music and fast visuals that evidently delivers the liberation. The piece continues following similar trajectories.

5.5.3.3 Liberation

Multiplexes that underpin liberation display a maximal degree of temporal freedom between sound and image available inside the framework of the isomorphic paradigm. As discussed in 3.3.4, "The Compositional Functions of Macro-isomorphism", this can be achieved by combining macro-isomorphism with music video style montage. Examples are the "climax" section of *Capillaries Capillaries* (16:15 – 17:33) or the "climax" section of *Singing Sand* (9:51 – 11:53).

These *multiplexes* function as the liberation of sound and image from an object, and thus the liberation from the constraints of continuous time and space. The visual objects or clusters point towards feelings and associations induced by energetic music. The time windows used are relatively short and the visual motion is relatively fast in order for the visuals to support the musical energy.

Since such *multiplexes* represent the final stage of *escapology*, I try to postpone the point of liberation in my pieces for as long as I can.

5.5.3.4 Quasi-Liberation

Quasi liberation refers to scenarios where the ties between sound and image are loosened but there is no sense of climax typical for liberation. The sense of audiovisual tension is hence still present. Quasi-liberation can be achieved by either "loosening" a *monoid* or "tightening" a *multiplex*.

"Loosening" a *Monoid*

An example of "loosening" a *monoid* can be found in *Capillaries Capillaries* (1:00 – 1:30) where a set of points and lines get sucked into a whirlpool. The first stage of this process starts as a tight *monoid*, in which there is a cross-modal correspondence between the energy of spinning and the pitch of an auditory stream. The rising of energy is gradually neutralised by slowly lowering the pitch whilst continuing to increase the visual energy by increasing the speed of the spinning. A strong cross-modal correspondence consequentially turns into a week correspondence creating a sort of cognitive divergence typical for loose *monoids*. At the final stage before the explosion at 1:30, the *monoid* is "loosened" to the extent of loose macro-isomorphism which results in a dreamy or associative audiovisual qualities detached from quasi-concrete reality. My compositional aim here was to make the upcoming explosion at 1:30 function as a surprise.

I used a similar technique when creating visuals for *Synspecies – Spaceless Latitudes* between 4:20 and 5:18. Here a visual object is slowly moving away from the viewer and fading into the distance. Simultaneously, a musical drone moves in the opposite direction by becoming louder and brighter. As in the previous example this creates a scene with dreamy or more poetic qualities typical for loose *monoids* (due to cognitive divergence). As a new visual object enters the scene at 5:18 "from behind" the viewer, a sonically anticipated moment finally gets its visual confirmation and reveals the future direction of the piece. The aim here was again to play with the expectations of the audience.

"Tightening" a *Multiplex*

Multiplexes can be "tightened up" by adding micro-isomorphic elements. Many examples can be found in *Singing Sand*, where *multiplexes* made from techno music and unrelated visual segments, are combined with micro-isomorphic elements. The micro-isomorphism is derived from sonified geometry. An example of such a *multiplex* can be experienced towards the end of the section between 2:30 and 4:06. In this case, the *multiplex* is an audiovisual loop (4 repeating visual segments accompanied by an audio loop). The visuals, which are also used to generate audio which gets gradually mixed in towards the end of the section. That "tightens" the *multiplex* and creates a tension with qualities typical for audiovisual objects (an associative space of *multiplexes* gets reduced to the spatio-temporal limitations of audiovisual objects).

Tight *multiplexes* can be observed towards the end of *Synspecies – Spaceless Latitudes* (6:02 – 6:14). The visuals in the last section consist of three windows that represent different perspectives and variations of a single visual object. Towards the end of the section the visuals start to flicker at the rate of a pulsating sound. This micro isomorphic element "tightens" the *multiplex*.

6 Audiovisual System & Workflow

In this chapter I will discuss my workflow and present the infrastructure of my audiovisual system. The infrastructure will be described from a composer's perspective, and hence I will not go into technical details. In the following chapter, "Details of Technical Implementation", the most important technical challenges behind the infrastructure will be discussed.

The aim of this chapter is to show that the aesthetics I have developed derive from a workflow that can be summed up as dialogue between ideas and technology. My role in this interaction is being a composer, creative coder and technician. In my case all these roles feed into one another making the distinction between them unclear.

6.1 Approach

The audiovisual system I developed for my own compositional purposes integrates the *Max/Jitter* patching paradigm for creating real-time *openGL* 3D graphics with the *Ableton Live/Max for Live* DAW paradigm. Such a system allows me to work simultaneously with sound and image and combines the flexibility and strengths of these two worlds. The Majority of audio synthesis, sampling and processing takes place in *Live* whilst *Max* takes care of geometry synthesis, sampling, processing and rendering. In cases where the visuals are sonified, the sound is synthesized in *Max* and then processed further in *Live*. The audiovisual system is bi-directional meaning that audio signals, parameters and events can be sent between audio and visual domains in either direction. This makes the system appropriate for working with any of the defined hierarchies between sound and image (see 5.1 for hierarchies).

Live functions as a *DAW* and the control centre of the audiovisual system. The timeline in *Live* provides a stable time, as opposed to frame-rate dependant time in my *Max* patches[22], hence all the fixed compositional events are defined here – either as automation data, MIDI notes or audio clips. These three elements, but mainly automation data, are used to control visual parameters in *Max* as well as to turn on and off various pseudo-random events in *Max.* These pseudo-random events can be recorded back into *Live* as automation

[22] Time in my *Max* patches is not stable as it depends on the frame-rate which is determined by the *qmetro* (queue-based metronome).

data, edited and then played back. Similarly, when I use hand gestures to interact with geometry via a *Leap Motion* controller, I can record the parameters of my hand gestures in *Live*, edit them if necessary and play them back – as one would edit and playback the recorded automation from a musical interface. The controls for playing back geometry are also in *Live* – I use automation curves for that purpose.

Whilst *Live* functions as a "compositional GUI" and hence my main compositional environment, *Max* represents a lower-level compositional environment. On one hand, it is a modular system for working with geometry whilst on the other it connects the whole system together by linking the modules with *Live*. This modular system consists of numerous modules that can be connected as desired to achieve a particular type of processing. The modules are not fixed, but rather in a stage of constant development. During the process of composing I am constantly improving them by adding new features, adopting old features and fixing bugs. If necessary I might also implement new modules. These things depend on demands of the composition. However, I try to patch as little as possible when composing.

Capillaries Capillaries was the first piece that utilised the audiovisual system outlined here. Therefore, the system evolved along with the piece. One of initial problems I faced was that I lacked the appropriate compositional experiences in order to know what features would I need in my system. On the other hand, I could not gain the necessary experience without having a system I could work with. Thus, I started the compositional process of *Capillaries Capillaries* with a loose plan, only because I had to start somewhere, as I did not really know where to begin. As I was developing the piece I faced numerous technical as well as compositional challenges and very often I had to take a few steps back and re-do things differently in order to continue. There were many compromises I had to accept due to hardware or software limitations, or due to my limited skills combined with time limitations. Part of my initial workflow was hence to constantly move forwards and backwards in technological and compositional sense due to either technical limitations or due to inappropriate working methods. This was the most stressful period of my PhD research, especially as *Capillaries Capillaries* grew into a longer piece, and all the way until the end I could not know what aesthetic compromises I will have to accept due to technical problems and limited amount of time. The solutions to the mayor issues I faced are described in chapter 7.

Currently my workflow consists of constant switching between:

- planning and intuition.

- experimenting with tools to generate compositional ideas and implementing tool for realisation of ideas

- creative and technical programming.

- mixing, sound design, visual design etc.

6.2 Infrastructure of my Audiovisual System – Overview

In this section I will present different families of modules and applications that together with *Max* and *Live* represent my audiovisual system. A module refers to either a *Max abstraction* or *Max for Live device*, whose architecture includes inputs and outputs, while an application refers to a separate *Max patch* without connectable inputs or outputs. The main difference between modules and applications is that modules can be connected together in a wider network of modules, like modules in an audio modular system, while applications cannot be integrated into the network of modules. Also, modules can be controlled via automation in *Live* while applications cannot.

6.3 Applications, Modules and Workflow

In this section I will present different applications and modules that I use in my work in an order that reflects my workflow.

6.3.1 Creating a 3D Model – 3rd Party Applications

In order to use a 3D model I need to first create one. An example of the use of the 3D model in my work can be seen in *Capillaries Capillaries*, where all the visuals come from a processed 3D model. Different segment of the same model can be also seen in *Synspecies - Spaceless Latitudes* (2:00 – 3:16; 5:17 – 6:14) and *Synspecies – Teaser* (0:38 – 0:47).

This original 3D model was created with the *Autodesk's 123D Catch* application that enables a user to create 3D models from photos of concrete objects. The photos of the object need to be taken from many different perspectives. *123D Catch* is unfortunately discontinued.

3D models generated using *123D Catch* need to be trimmed as there are usually more elements contained in the model than desired (an example would be a floor on which the object was positioned). For trimming and additional mesh repair I use *Autodesk's Meshmixer*. *Meshmixer* offers numerous features such as mesh modifications, vertex reductions and model groupings. Once the 3D object is prepared to be exported as an *obj* file, which I can import in *Max*, I check if the number of vertices correspond with an estimated number of vertices that I assume my computer would be able to process in real-time. If there are too many vertices I trim portions of an object or reduce the number of vertices – often both steps are necessary. Technical limitations have therefore aesthetic implications.

6.3.2 Application for Optimising the 3D Model for Working Within *Max*

After importing the 3D model into *Max* using the *jit.gl.model* object the number of vertices in the output matrix increases by, on average, six times the number of vertices in the original model due to the method that *jit.gl.model* uses to draw the mesh. The reasons behind this are described in section 7.4.1. This presents an additional CPU utilisation, hence, I programmed an application which reduces the number of vertices back to the original number. This process needs to be carried out only once, and after it is finished the 3D model can be saved into the *Jitter matrix* file format.

6.3.3 Geometry Synthesis, Geometry Recording and Geometry Playback

The source for geometry synthesis can be geometric noise (randomly placed points in space), a basic shape (sphere, cube, plane, cone etc.) or any other 3D graphics with at least x-y-z position data. As in case of sound synthesis, the source is then further processed[23]. The processing is done by the same modules that I use for geometry processing (see section 6.3.4), although the architecture in synthesis applications is fixed. Once I implement a geometry synthesizer I tend to not change it at least in the course of composing a piece. The inspiration for using such approach comes from hardware synths with a limited number of oscillators, envelopes, LFOs etc. Working with limited possibilities forces me to explore the aesthetics of these limited possibilities. When synthesizing geometry, I often aim for very fragile states for each object, states that are kept alive by very precise settings of parameters and are almost impossible to control (see *Singing Sand* 5:41–5:55; 7:05–7:12; 5:20–5:29). That is a main difference between the material I synthesize in advance and then sample when composing, and the material I generate during the compositional process. The latter is more stable and more controllable. An example of the latter can be seen in *Capillaries Capillaries* (14:12-14:45)

Geometric gestures I synthesize with applications are not controlled by automation. Using an analogy from the audio world, the gestures are like "one shots" that get reset each time they are triggered. Due to lack of control I always record the output of the geometry synthesis application as x-y-z data and store them on my computer as one would store audio samples. When composing I play the visual gestures back using a geometry playback module that supports also time-stretching. Technical details of geometry recording application and the geometry playback module are discussed in section 7.2.2.

6.3.4 Composing with the Timeline and Geometry Processing Modules

In order to start composing with the timeline, I need to:

- in *Live*, instantiate a *Max for Live* device that sends *Live*'s automation data to *Max* (see Figure 4)

- in *Max*, instantiate either a geometry playback device (see Figure 5) or import a converted 3D model into a *Jitter matrix* (see section 7.4.1 for detail on conversion)

[23] When processing 3D object I always process only position data and leave other planes of the matrix intact.

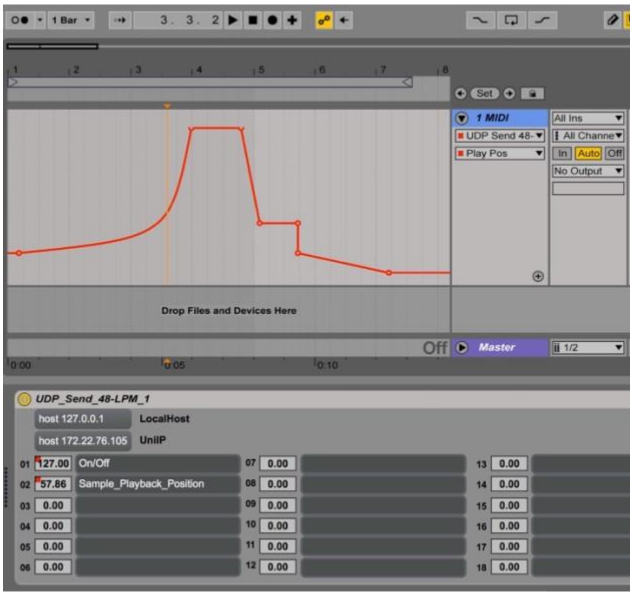

Figure 4: The beginning of the compositional process for the piece *Singing Sand* in *Live*. The *Max for Live Device* that sends automation data from *Live* to *Max* with 48 *live.numbox-es* (only 16 of them are visible). The *live.numbox* with number 01 sends automation to UDP port 8501, 02 to 8502 etc. The two automated parameters control the geometry sample playback position and switch the module on or off.

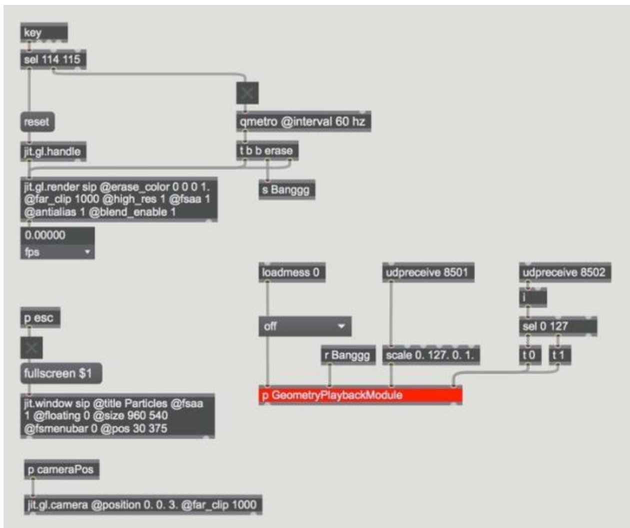

Figure 5: The beginning of the compositional process for the piece *Singing Sand* in *Max*. The *Max* patch consists of a geometry playback module, two UDP ports that receive automation data from *Live* and other *Jitter* objects necessary for establishing and displaying a rendering context.

Once the geometric sample is inside *Max* I use automation curves in *Live* to control it. For each new automation curve that I use I need to add an UDP receive port inside *Max*. The output of the receiving UDP port needs to be scaled and connected to the desired parameters (see Figure 4 and Figure 5). I add names of the parameters to the *Max for Live Device* as I compose.

In *Max*, I always start with the modules that are necessary to establish what I call an embryonic audiovisual object – that is the object that I develop before I start composing with the timeline. The embryonic audiovisual object determines the essential qualities of the piece (see section 5.2.2 for more information on embryonic audiovisual object). In the course of the compositional process, the system of modules grows along with the piece. The final configuration of modules for two parts of *Capillaries Capillaries* can be seen in Figure 6 and Figure 7. For the whole piece I used four different configurations of modules, or in other words, four different *Max* patches.

Common to all geometry processing modules is that they do not have an on/off switch, but rather they are turned off by setting the amount parameter (or equivalent) to 0. That way I can control the amount and on/off state with one parameter. When the module is turned off the data flow is passed by instead of being processed. That enables me to save CPU power when building large and complex patches like the ones presented in Figure 6 and Figure 7.

Live serves as a fixed linear timeline, while the frame rate in *Max* is driven by *qmetro* object which skips the frames it cannot process in time. Consequentially, all the fixed and most important temporal events are specified in *Live*. *Max* on the other hand works in a "fluctuating time", due to the fluctuating frame rate of updating and rendering a real-time simulation. *Max* also handles the pseudo-random events or sets of pseudo-random events, whose start and end are determined in *Live*.

One of the most important things in my workflow is that when I shape my audiovisual material, I can simultaneously watch it and listen to it. Whenever I press *play* in *Live*, I hear the audio and see the visuals. However, if the processing of the visuals engages physics-based algorithms, then I need to start with the playback somewhere in the timeline prior to the point at which they switch on.

At the point when the modular system in *Max* is set and *Live*'s automation parameters are mapped to visual parameters, I compose solely from *Live*, drawing automation curves to control both sonic and visual elements. In case I also use *Max for Live*'s envelope followers or built-in transient detection to additionally control visuals, I convert the analysis data into editable automation curves in order to obtain greater control (see section 7.3 for more information).

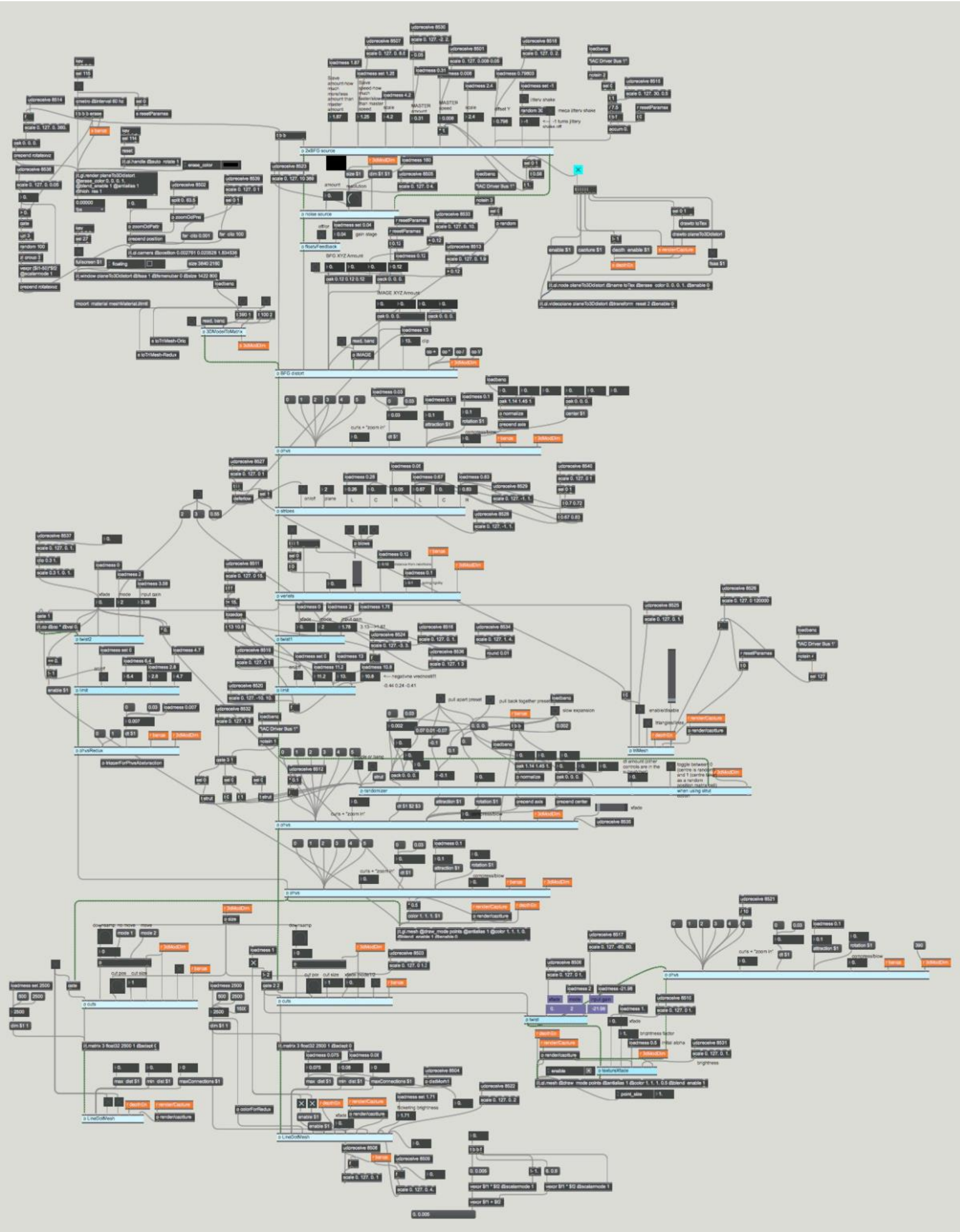

Figure 6: The configuration of geometry processing modules used in *Capillaries Capillaries* for the part between 11:40 and 17:33. Each module has a blue colour.

Figure 7: The configuration of geometry processing modules used in *Capillaries Capillaries* for the part between 17:33 and 19:40. Each module has a blue colour.

6.3.5 Deforming Time with the Stringvencer Application

Stringvencer is a custom-made sequencer that enables deformations of pre-composed temporal structures. These temporal structures are expressed as MIDI notes and the application uses physics-based algorithms to displace these MIDI notes in time. *Stringvencer* is a CPU intensive patch as it runs in high priority with high refresh rates in order to deliver high temporal resolution. This is the reason I treat it as an application and I do not run it alongside other CPU intensive applications. For technical details about *Stringvencer* please see 7.1.

In order to use *Stringvencer*, a MIDI sequence is first composed in *Live*. In some cases, the sequence is derived from an audio track using *Live*'s *slice to new MIDI track* function. *Stringvencer* allows the user to import up to ten separate MIDI sequences. If more sequences are needed, two or more can be merged into one inside *Live* by mapping MIDI notes parameters to different instruments. The maximum length of a MIDI sequence supported by *Stringvencer* is thirty-two bars. When my aim is to distort the musical time slowly across a longer period, the MIDI sequence is used in a loop. That was the case in *Capillaries Capillaries,* in which I composed an eight-bar long audiovisual loop and then distorted it slowly over time – first by distorting the musical part and then by separately matching the visual part. The original audiovisual loop can be heard at 11:41, followed by the gradual temporal distortion.

Once the MIDI sequences are prepared, each separate sequence (each MIDI clip from *Live*) needs to be converted into a *Jitter matrix*. This is done using a custom-made *Max for Live Device* that executes the conversion. The *Jitter matrix* consisting of MIDI data can then be imported into *Stringvencer*. Once *Stringvencer* is running (looping through the imported sequence) the playback should be identical to the original playback in *Live* (assuming that *Stringvencer* sends MIDI data back to *Live* and triggers the same instruments). At this point, one can interact with *Stringvencer* whilst the temporal deformations can be recorded as MIDI notes back into *Live* (see *Stringvencer Tutorial* video for more information).

Once the MIDI notes are back in *Live*, I need to either match the visuals with the displaced notes or compose a new visual sequence on top. I do this using automation curves, which need to be tightly synchronised with the MIDI notes. Since the deformed MIDI notes are off the grid, I have to use horizontal zoom to accurately position the break points on the automation line. This process can be time consuming, hence, I made a *Max for Live Device* that positions the break points on an automation line according to the onsets of MIDI notes (see Figure 8). As a result I have to adjust only slopes and curves which allows me to concentrate only on the creative aspect.

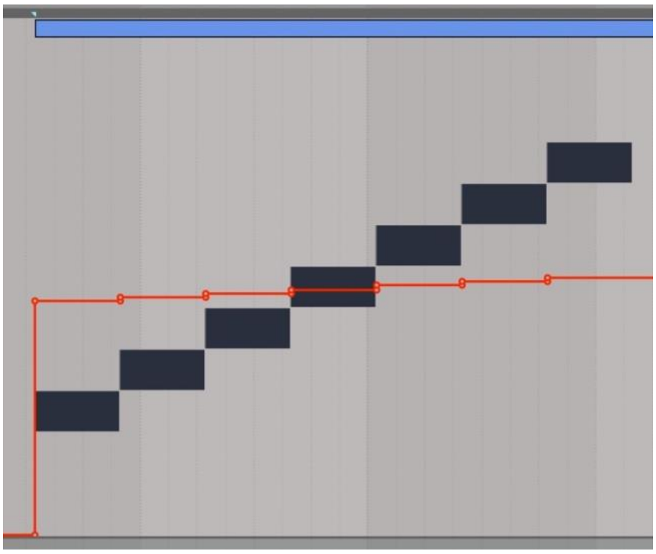

Figure 8: Using a custom-made *Max for Live Device* to convert the onsets of MIDI notes to break points of an automation line.

6.3.6 Offline Render: Automation Recording and Reading

Once I finish composing on *Live's* timeline I render the sound and image in offline mode in order to obtain the maximum possible visual quality. All the submitted compositions exist as fixed media pieces but *Capillaries Capillaries* and *Singing Sand* also exist in a performance format. *Capillaries Capillaries* was rendered offline in four parts and I used four different *Max* patches to compose the whole piece. *Singing Sand*, on the other hand, was entirely composed using only one *Max* patch. Therefore, I can play back the whole piece with real-time visuals which gives me a lot of flexibility in live performances[24]. The latter approach, one *Max* patch and one *Live* project, represents my ideal compositional scenario as it enables me to keep everything open to change until the end of the compositional process – excluding creative editing in post-production which I sometimes do (see section 6.3.7 on post-production). Hence, I try to compose as much as I can with one *Max* patch before I render the visuals offline.

[24] It would be possible to merge the four patches from *Capillaries Capillaries* and play the whole piece back in real-time but I've decided to use video clips when performing and focus predominantly on the audio aspect.

The first step of the offline rendering process is capturing the rendering context into a texture (to move from vector graphics to raster graphics). This requires adjusting the sizes of points, widths of the lines and opacity of all the drawing primitives in the piece for each separate visual scene as the captured version looks very different (see section 7.4.2 for the visual comparison). The technical reasons behind this problem are not know to me and according to the correspondence with *Cycling '74* there is not instant solution to this problem. Therefore, I have to draw the automation steps for these new values in *Live*. Once the captured scene looks as it should, the automation data from *Live* needs to be transferred into a *Jitter matrix*. I do this by recording the automation data into matrix in real-time using a custom-made application (see section 7.4.2 for more details). Once the automation data is transferred, this automation data matrix is saved to disk. The next step is to add a module for reading automation data, into the *Max* patch, into which the matrix is loaded. When the matrix is loaded, the functionality of *Live*'s timeline in regard to the visuals is replicated in *Max*. This enables me to move in time across the whole piece inside *Max*. At this stage everything is prepared for offline rendering.

During the offline rendering process, *Max* generates all the MIDI events derived from the visual motion, that will be used to trigger samplers or synths in *Live*. This MIDI data needs to be captured and converted into MIDI clips inside *Live* using a custom-made *Max for Live Device.* This process is described in more detail in section 7.4.2. Once the MIDI clips are in *Live*, audio can be separately render offline. The audio and the video file are then joined and synchronised inside a video editing software.

6.3.7 Post-production

The very last stage of my workflow is the post production. Post production consists of three parts: finalising audio (without an image), finalising video (with whatever audio mix is available at the time) and final corrections on the audio and video while listening and watching both mediums together. I also send the video to the mastering engineer with a request to either preserve or emphasise certain audio qualities that are important as part of the audiovisual whole.

By finalising audio I mean creating a final mix while complying with my compositional aims. If, for instance, my compositional aim is to create a strong contrast between two sections, I amplify this contrast by employing mixing elements such as quieter - louder, narrow stereo – wide stereo, muffled – brighter, contrasting acoustics simulated with reverbs and delays etc.

The video part of post-production usually consists of corrective editing and subtle addition of post-production effects. I use *Final Cut Pro* to edit visual glitches that occur due to imperfections in *Max* patches or imperfections while using automation curves in *Live*. An example of the latter case would be a situation, where two or more parameters need to be changed simultaneously, but since I mostly work without the temporal grid the bulk changes might not be perfectly synchronised. Apart from this, I mostly use *Final Cut Pro* to do exposure or colour corrections and to apply some blur where needed. I set the parameters of these effects for each separate shot, and hence I chop the long video clips I get from *Max* into many short clips. However, section one and four in *Synspecies - Spaceless Latitudes* (0:00 - 2:00; 5:17 – 6:14) are examples in my work where I used video editing software for compositional purposes. For section four, for instance, I recorded a single visual object from different angles and applied different types of geometric distortion to it. Recorded takes were then edited in *Final Cut Pro*.

When mixing the music in the studio I tend to make the mixes too "soft" in regard to the audiovisual whole. I try to fix such things at the end when correcting the mix according to the image and vice versa. Sometimes, for example, the "aggressive" or the "burned" qualities of the image demand harsher and brighter sound. The amount of timbral harshness and brightness that would often fit best can be too ear fatiguing, especially when considering loud volumes in average performance spaces. Hence, at this last stage of finalizing the audiovisual whole, I am always trying to find a compromise between demands of the sound and demands of the image.

6.4 Aesthetics as a Dialogue Between Technology and Ideas

Here I will sum up the global aesthetic implications of my workflow.

I believe that my workflow and the audiovisual system I implemented reflect my wide range of interests into different aesthetics, which I attempt to combine in my work. The musical influences span between techno, experimental electronic music, free improvised music, ambient, glitch or acousmatic. whilst the visual influences come from physics-based algorithmic visuals, 3D wireframes or point-clouds. In order to operate in between these various aesthetic categories, I have to constantly switch and juxtapose various techniques of composing, production, mixing or programming. Hence, my audiovisual system, needs to be very flexible and I think it is this flexibility that makes it different from other existing tools for the audiovisual production. For instance, sometimes I interact with my real-time audiovisual material using a *Leap Motion* controller, and the resulting aesthetics are reminiscent of experimental computer improvisation, while on the other hand my tool offers relatively hi-quality offline rendering, which is usually not associated with the improv aesthetics. The quality of offline rendering is, however, still far from what industry standard renderers offer but that gives the visual aesthetic a different twist that I personally prefer in this context. In the same way as the visual aesthetics span somewhere between "lo-fi" and "hi-fi", so is the sound world, where I compose using "electromagnetic" sounds recorded with a DIY microphone that cost me three pounds to build, along with high-quality samples and expensive plug-ins. Also, both sound and visuals are sometimes carefully designed while sometimes controlled by virtual forces and pseudo-random events. The system, the workflow and the overall aesthetic in my work hence all feed into one another and all of them need to be considered when discussing each element in isolation.

7 Details of Technical Implementation

In this chapter I will talk about the most important technical problems that I had to solve in order to implement and refine my audiovisual system, which operates between *Max* and *Live*. The focus of this chapter will be purely on the technical side of that implementation.

7.1 Stringvencer Implementation

Before reading this section, one should be familiar with the workflow of using *Stringvencer*. Please refer to *Stringvencer Tutorial* video.

Stringvencer opertes between *Max* and *Live*. *Live* handles the original MIDI sequences[25] and all the audio synthesis and sampling. The MIDI sequences from *Live* are copied into a *Jitter matrix* which enables the MIDI sequencing to be moved from Live into *Max*. Therefore, *Max* takes care of the MIDI sequencing (and time distortion) that triggers synths or samplers in *Live*.

The copying of MIDI data from *Live* into a *Jitter matrix* (I will call it a MIDI matrix), is done by *Max for Live Device* I named *NotesToJitterMatrix*. This device copies the pitch, position, duration, velocity and on/off state from *Live* MIDI clips into separate planes of a matrix. Each matrix cell, with its 5 planes, hence represents a MIDI note. The note duration for a semiquaver in *Live* is 0.25 and the same goes for position. A quaver at the beginning of bar 2 would hence have a duration of 0.5 and position 4, as the counting of beats starts with 0. In *Max* the only plane of the MIDI matrix that gets processed is the position plane. All other planes remain intact.

The string representing the timeline is actually made from 1 rigid body per 1 semiquaver, hence the length of the MIDI sequence needs to be specified when importing the MIDI matrix. The rigid bodies, visualised as points, are linked by springs and their movement has six degrees of freedom – all that is taken care of by *jit.phys.multiple*. I use a system of rigid bodies to simulate a soft body as *Max* does not support the simulation of soft bodies. This semiquaver grid of rigid bodies, which I will call simply points, serve as a reference for the position of MIDI notes regardless of their position on the timeline. For instance, when the position of a MIDI note, as specified in *Live*, is 0,125 (demisemiquaver

[25] MIDI sequences can be generated also from audio tracks by using *Live*'s 'slice to new MIDI track function' which chops the audio file and moves the audio chunks into samplers.

from the beginning), it means that it sits exactly in the middle between the first and the second point. When the string gets deformed, this ratio (50% of semiquaver from point one and 50% of semiquaver from point two) is preserved – no matter how deformed the string is, the MIDI note will always stay exactly in the middle between points one and two. The position of MIDI notes in the original MIDI sequence however does not have to be on any grid. If a note is for instance 3% of a semiquaver away from point one and hence 97% of a semiquaver away from point two, the ratio will be preserved when the string is deformed.

As a result of the approximation of a string as a set of solid bodies, one of the first steps that *Stringvencer* performs when importing a MIDI matrix is that is calculates the offset of the position of each MIDI note from the underlying grid of the string. The offset from the closest neighbouring points on both sides is then expressed as a relative measure and preserved while the string is deformed.

The temporal resolution of *Stringvencer* is however not limited by the rounding of the numbers that represent position and duration, but with the frame rate of the physics engine. In *Stringvencer* the automatic refresh rate of the physics engine is turned off in order to be driven by the metro object that runs in high priority. The refresh rate of the high priority scheduled events is limited to 1ms in *Max*. In practice, with my 2013 Macbook Pro[26], the refresh rate setting of approximately 500Hz worked well for processing a single MIDI sequence while the refresh rate had to be lowered to around 200Hz when processing approximately 8 MIDI sequences simultaneously. All the processes that affect the sequencing of audio events are updated with such high refresh rates and are hence separated from the part of the patch that takes care of visual rendering.

In order to achieve a good performance on my computer, the geometry (points) needed to be processed by a shader that does not use fixed function shading and that is why the material is applied to the geometry. This seems counterintuitive since additional shading represents an extra burden for a graphics card. As it turned out, the basic graphics in *Max* are processed by fixed-function *openGL* pipeline while many newer cards (including NVIDIA GeForce GT 750M in my Macbook pro) are not optimised to perform fixed-function shading. Without using a shader that uses programmable pipeline, the interaction with the string was practically impossible due to drastic fluctuations in the frame rate of the video rendering.

[26] MacBook Pro (Retina, 15-inch, Late 2013), 2.6 GHz Intel Core i7, 16 GB 1600 MHz DDR3, NVIDIA GeForce GT 750M 2048 MB

7.2 Implementation of Specific Visual Algorithms

7.2.1 Connecting the Points

I implemented the algorithm that connects points with lines according to the distance between the points first in *Max* (see *Max* patch *Connect Points - Max Version*), then in *JavaScript* (see *Max* patch *Connect Points - JavaScript Version*) and at the end in *Java* (see *Max* patch *Connect Points - Java Version*). The algorithm is not optimised so it calculates the distance between each point and all the others. With implementation in *Max* I could run approximately 120 points at 60 fps on my late *2013 Macbook Pro*. I was drawing points with *jit.gl.mesh* and lines with *jit.gl.sketch*. I used the same approach with the *JavaScript* version, although instantiating *jit.gl.sketch* within the *JavaScript* external. This implementation enabled me to run approximately 500 points at 60 FPS. The Java implementation, on the other hand, enabled me to run around 2000 points at 60 FPS, which was enough to create appealing dynamic textures. This version does not rely on *jit.gl.sketch* but creates an index matrix for *jit.gl.mesh*.

A thing that prevents the *JavaScript* version from working faster is the fact that there is only one method available to grab data from a *Jitter matrix* and that is to use the *getcell* method. The *getcell* method needs to address each matrix cell separately, hence, the bridge between the C world of *Max* and *JavaScript* needs to be crossed as many times as there are cells in a matrix – and these calls are CPU-intensive in such high quantities. The advantage of Java on the other hand is that it can pull the whole matrix into Java land at once and then all the individual cell comparisons can be done within Java.

7.2.2 Recording and Playing Back the Geometric Movement

There are a few different possible approaches to recording geometric movement (x-y-z data). The method I use records the x-y-z position data of vertices for each frame as a separate *Jitter matrix* file on a disk. The recording can be done in real-time or offline. The implementation of this method is very simple – for each frame a "write" message gets sent to the matrix that performs recording. No indexing or naming is need as *Max* automatically indexes the recorded files in order to avoid duplications - instead of one video file I end up with hundreds or thousands of files inside one folder.

For playback I use a frame-interpolation technique that enables me to slowly fade between two successive recorded frames (two successive frames are loaded into two *jit.matrix* objects). That enables me to time-stretch recorded geometric gestures without lowering the output frame rate. Frame-interpolation time stretching works very well for x-y-z data as the positions of vertices simply morphs between the two frames. The time-stretching can, however, look artificial if the position difference between two frames is too large. In order to avoid this issue higher recording frame rates are needed. However, sometimes time-stretching artefacts which are essentially linear x-y-z interpolations, can function very well inside a composition as the rigid movement, at least in my case, is very contrasting to the otherwise organic gesturality.

7.3 Audiovisual System Infrastructure

The infrastructure of my audiovisual system includes *Live*, *Max for Live* and *Max*. *Live* handles all the audio management (synthesis, sampling and processing) and all the sounds that originate in *Max* are routed to *Live* via either *Soundflower* (Ingalls, 2014) or the *Loopback* (Rogue Amoeba, 2016) application – I use Loopback when I need to route audio in both directions. While Loopback is very flexible, its latency depends on the sampling rate, resulting in a latency of 25ms at 44.1kHz according to the manufacturer[27]. Sending a signal from *Live* to *Max* and back is hence already problematic. The Jack Pilot application, on the other hand, behaves inconsistently on my system so I stopped using it. *Max* takes care of geometry processing, any sound synthesis and triggering of MIDI notes that originate in geometry. Visuals are, of course, also rendered in *Max* using *Jitter*.

Importantly, *Live* serves as a timeline for controlling *Max*. At first I was using MIDI notes from *Live* to trigger events in *Max* and automation in *Live* only for continuous controls, but later on I replaced MIDI notes with automations. In cases where I generate some MIDI notes with MIDI devices in *Live* that I want to route into *Max*, I convert the desired MIDI information into automation with a custom-made *Max for Live Device*. The converted MIDI note data looks like a step function. The advantage of automation data is that one can move either continuously, linearly or exponentially between two steps if needed. In a similar manner, I convert the results of audio analysis tools from *Live* into automation data. Audio analysis might be performed in real-time, for example using an envelope follower, or offline, for instance using the built-in transient detection. To implement the first case I modified the envelope follower from *Max for Live* so it generates automation directly, which can be easily recorded. One way to implement the latter case would be to slice an audio track at the transient markers by using the 'slice to new MIDI track' method,

[27] This detail was revealed in an email corespondence with Chris Barajas from Rogue Amoeba.

and then convert the resulting MIDI data into automation data. Having all the controls in automation format enables a uniform mechanism for manual editing.

Automation data from *Live* is sent to *Max* via UDP messages. I tend to keep all the automations I use for controlling *Max* on one track in *Live* – hence all the *live.numbox* objects that correspond to *Live*'s automations reside in one *Max for Live Device*. This *Max for Live Device* holds 54 *live.numbox* objects and in case I need more automations than one device is enabling me, I add another track with another device. Reading automation data from more than one *Max for Live Device* can cause automatic disabling of certain automation parameters (the automation parameters turn from red into grey) and in order to prevent that all *live.numbox* objects need to have the *defer automation output* attribute enabled. This however does not guarantee that some parameters will not become automatically disabled when the project becomes very CPU demanding, causing excessive CPU loads. One modification that can help is to reduce the interval of the *update limit* attribute of the *live.numbox* objects and to reduce the overall refresh rate of *Max*'s GUI in *Max* preferences.

I send each automated parameter to *Max* via a separate UDP port rather than packing and unpacking the data, because the automation outputs the values only when changed. Hence, when composing, I set the update limit to match the framerate of the animation in *Max*.

7.4 Performance Issues

7.4.1 *jit.gl.model* and the Output Matrix

When working on a project that demands a bi-directional communication between sound and image on different levels, *Max* seems perfect tool for the job. However, when one wants to work with CPU or GPU expensive materials and techniques in *Max*, *Jitter* presents many obstacles. One obstacle is the way in which *Jitter* exploits the available CPU and GPU power of computers nowadays. There is no simple way to make use of multithreading with *Jitter*, as it is in newer versions of poly~ for processing audio signals. In theory, splitting a patch into two or more patches and running them in separate *Max* applications should help with the multithreading issue, but I was not able to improve the performance in any significant way – most of the CPU cores were unused while *Max* was not able to deliver a desired performance. There is also no support for newer optimised rendering API's such as *DirectX 12*, *Vulkan* or *Metal* and there seems to be no support for *Crossfire* or *SLI* technology that enables the use of multiple graphics cards. Additionally, the implementation of certain features in *Jitter* is not optimal. One example is the output matrix of *jit.gl.model*.

jit.gl.model is an object for importing 3D models into *Jitter* and it has the ability to output the model in matrix format for further processing. The first problem is that the resulting output matrix does not have any colour data, so when drawing the model from matrix format using *jit.gl.mesh* the colour data from the model is lost. The next problem is that the output matrix consists on average of approximately six times more vertices than necessary (see *Max* patch *Application for Optimising the 3D Model for Working Within Max*). The reason for this is that *jit.gl.model* uses a *GL_Triangles* method, which is an older *OpenGL* method for representing the mesh, in which each triangle demands a set of three vertices in order to be drawn. In the example from Figure 9 that means that vertex v5 would be duplicated four times.

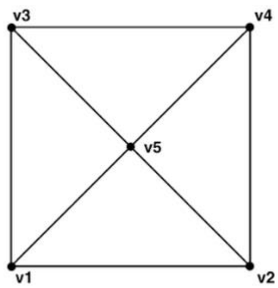

Figure 9: Vertices connected with lines

When a model consists of hundreds of thousands of vertices, this duplication of vertices results is a significant and unnecessary additional processing requirement. This issue is the reason that *openGL* offers a workflow that avoids this, using *index buffers* that use a minimum number of vertices and provide the set of instructions for the lines to be drawn between the vertices (Luten, n.d.). *jit.gl.mesh* luckily supports this option. Implementing this in *Max* requires a position matrix (x,y,z) and an index matrix (instructions on how to connect the vertices with lines so they will form a desired drawing primitive – in this case triangles).

In July 2016 I implemented an external in Java that eliminated all the vertex duplications from the position plane. That version generated a reduced position matrix but it did not create an index matrix and it also ignored the normals data from *jit.gl.model*. Hence the reduced matrix was appropriate only for use with the 'points' drawing. In 2018 I updated the external so it creates an index matrix for triangulation. Combining my external with *xray* externals, available through *Max*'s *packet manager*, I implemented the patch which creates a reduced matrix that includes all the available data from *jit.gl.model* (see *Max* patch *Application for Optimising the 3D Model for Working Within Max*). The update of my external was made with the help of *Max* forum user Greg Finger, who in the meantime implemented a solution in *Max* (Max Forum, 2016). The problem with Finger's *Max* version is that it is very slow. My very rough estimation is that it works approximately 1000 times slower. For larger models with hundreds of thousands of vertices, my external, which is not optimised, generates a reduced matrix in approximately 5 to 30 minutes. The *Max* version on the other hand would take days or weeks to perform the same task (the patch has a counter that counts the vertices it has already re-mapped so one can estimate the time needed for completion). Hence it is not useful for a compositional workflow where part of the experimentation process is to try out geometry processing for various 3D models.

7.4.2 Offline Rendering

During my compositional process all my compositions, or at least movements of the pieces, can be played back in real-time using transport controls in *Live*. The refresh rate of the real-time animation in *Max*, which is the same as the frame rate of the visuals rendering, fluctuates drastically due to occasional excessive CPU loads. Therefore, in order to avoid the noticeable frame-dropping in the recorder fixed-media versions and also to achieve the visual quality far beyond the quality achievable with real-time rendering, it was necessary for me to develop a number of patches and *Max for Live Devices* that enable me to render in offline mode.

Probably the most convenient way to use offline rendering in *Max* is to use the object *jit.record*. All my visuals are made using *Jitter openGL* objects and in order to record an *openGL* rendering context with *jit.record*, the rendering context needs to be captured into a texture. When rendering context consists of drawing primitives such as points and lines whose size and width respectively need to be specified in pixels, the captured texture looks very different to the original *openGL* 3D graphics when displayed directly in *jit.window* (see Figure 10 and Figure 11). This is one of the many unavoidable surprises when working with *Jitter* and it is impossible to make the two look the same. In order to come as close to the original as possible with the texture version, one needs to re-set the sizes, widths and brightness or opacity parameters for all the drawing primitives[28]. That creates a lot of extra work when working on longer pieces with different settings and automated parameters across the piece, as specific settings are needed for most of the 'shots'. Hence, additional parameters need to be manually automated across the whole piece.

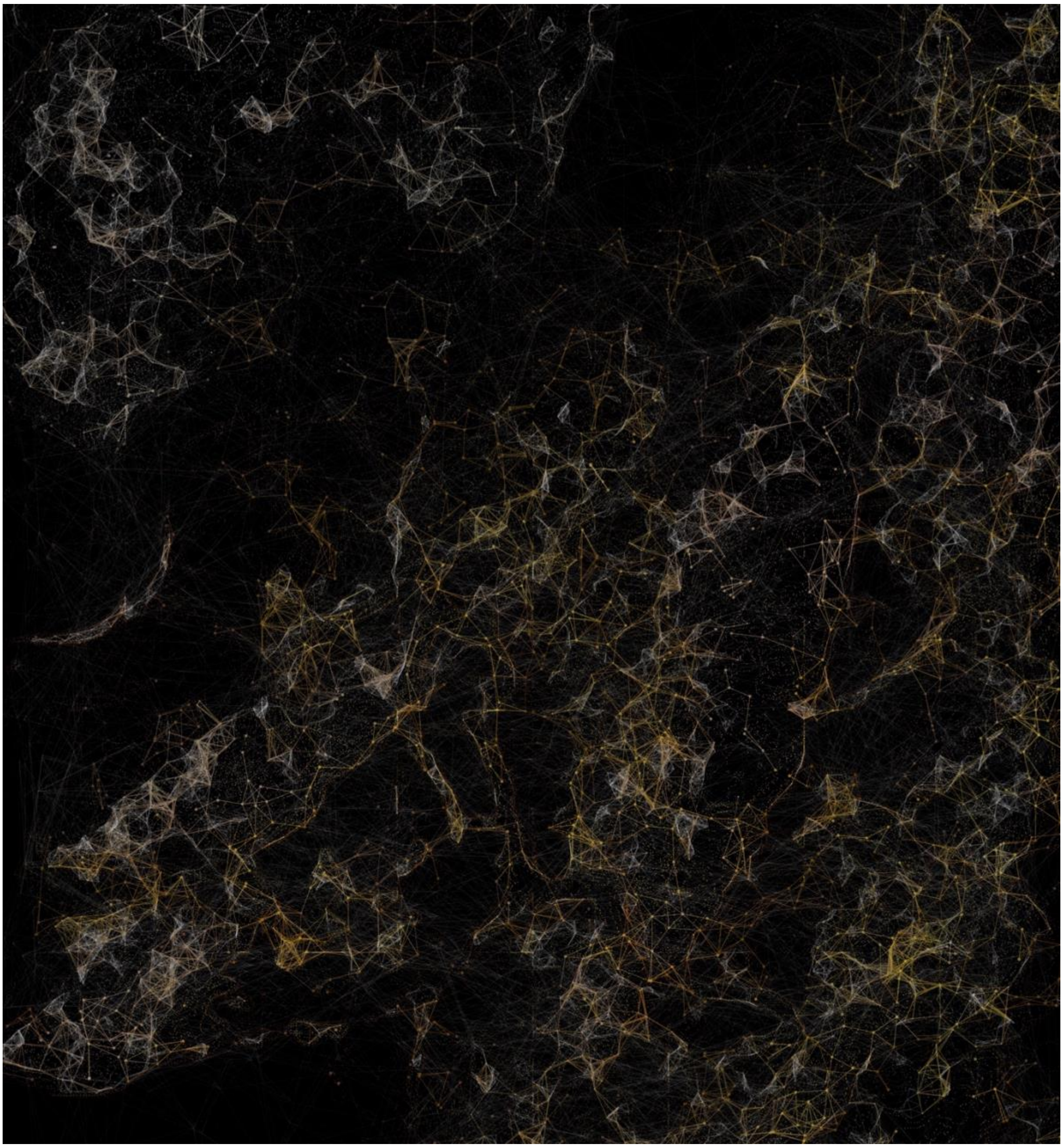

Figure 10: *OpenGL* rendering context consisting of points and lines and displayed in a *jit.window*. The rendering context is the same as in Figure 11.

[28] This solution was proposed by *Jitter* developer Rob Ramirez in en email corrspondence. The technical reasons behind this problem are unknown to me.

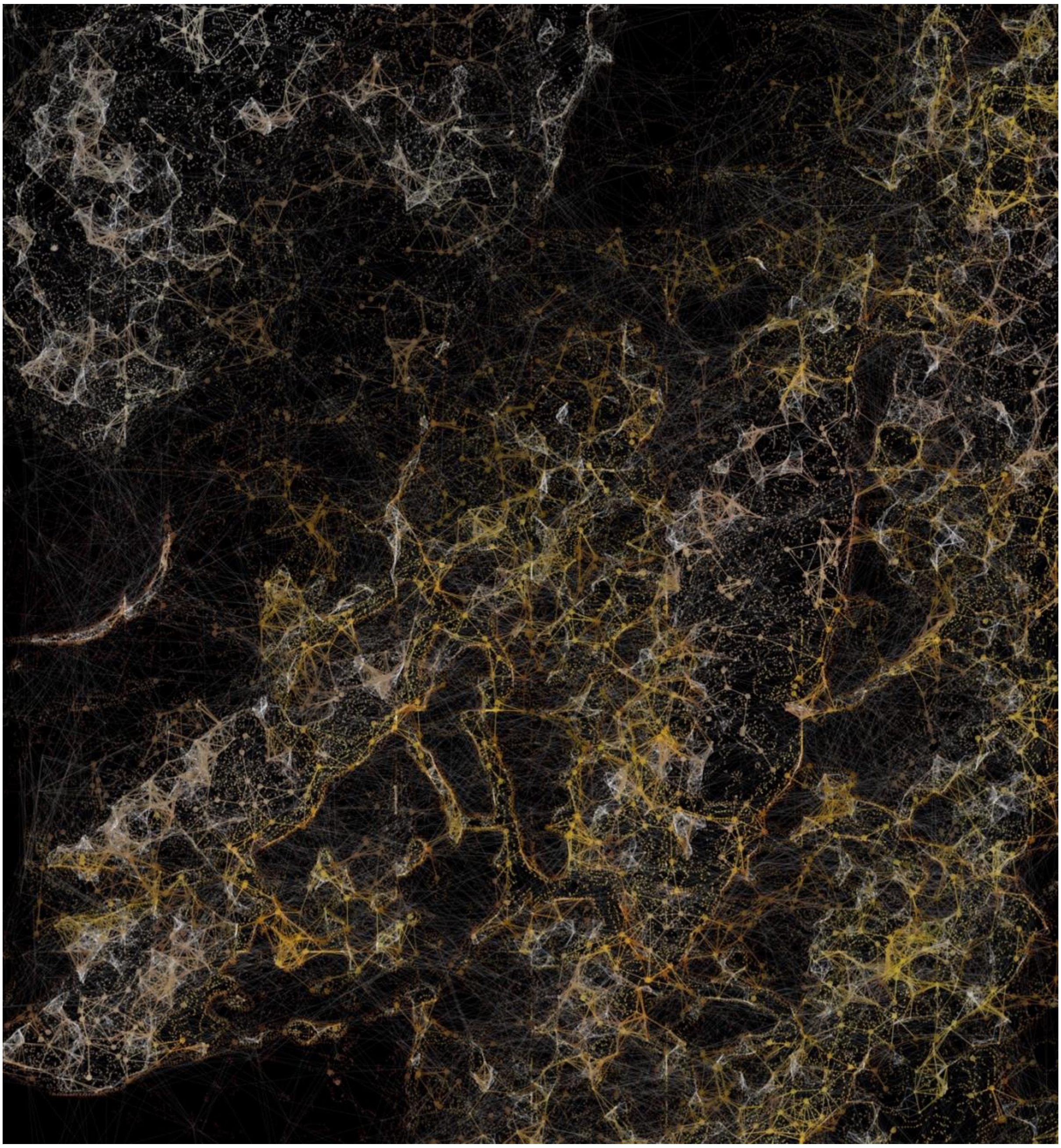

Figure 11: *OpenGL* rendering context consisting of points and lines, captured as a texture and displayed in a *jit.window* via *jit.gl.videoplane*. The rendering context is the same as in Figure 10.

Once the visuals are adopted to the texture format they can be recorded with *jit.record*. In theory, replacing the main *qmetro* in a patch with *uzi* and sending the desired number of *bangs* to the patch while recording should record all the frames. In practice, however, that works only for very simple patches. When working with more demanding material and larger patches this method can result in a crash. Therefore, I first attempted to record without using *uzi* by creating a feedback loop. The source of *bangs* in the feedback loop comes from *jit.record* which outputs a *bang* each time it records a frame - *jit.record* hence replaces the *qmetro*. The feedback loop is started by a manual *bang* using the *bang* object, while the looping can be stopped by adding a *gate* object to the loop. This approach also crashed *Max*. Adding a single *deferlow* object in the loop did not solve this problem either, but chaining three *deferlow* objects together did solve the problem. I have no idea why three *deferlow* objects are needed. This method of recording is very handy as one can count the frames and in case a specific fix is needed at a certain frame, such as adjusting the velocity parameter of a physically based animation, the frame count can be used as a trigger for appropriate further processing.

In order to render offline animation that is automated from *Live*'s timeline within a patch that also sends MIDI messages to *Live*, two additional steps are needed:

- All automations from *Live* (and MIDI data if any) that control *Max* need to be moved into *Jitter matrices*. This step should be done after all parameters are automated, including the correction factors for offline rendering.
- All MIDI messages from *Max* need to be converted into *Live* MIDI clips. This process takes place during the offline rendering.

In order to move automation data from *Live* into a *Jitter matrix*, I first remove all the tracks from Live except the one with automation data - since the automation data will be recorded into a *Jitter matrix* in real-time I want the best possible performance. Within *Max* I use a *metro* object to sample the incoming automation data while I set the *update limit* attribute of *live.numbox* objects to a small value (such as 2ms). If I leave the *update limit* attribute at the same rate as the framerate in *Max*, which is my preference during the compositional process, some automation updates can get dropped when recording. I record each automation on a separate Matrix plane so that each cell holds all the automation data for a given frame. The maximum number of planes in a matrix is 32, so more than one matrix must be used when there are more than 32 automation parameters. The *Max* patch that records the automation and the *Max* patch that reads the automation use the same UDP ports as I always use when controlling *Max* with *Live* (from 8500 onwards). Therefore, when integrating the *Max* patch that reads the automation into my parent compositional *Max* patch, I can simply copy-paste it in without having to do any extra patching.

While rendering the video in offline mode, all the MIDI messages that are generated inside *Max* are written as text data in *coll* objects and later saved as text files. Also, the frame at which each MIDI note starts and ends needs to be translated into position and duration values that *Live* understands. The parameters of each separate MIDI note together with its start and off times form one line of MIDI data within the text file. After the offline rendering process is finished the *coll* objects need to be imported in a custom-made *Max for Live Device* that converts the data stored inside each *coll* object into a *MIDI* clip. The largest MIDI clip that was generated for *Capillaries Capillaries*, for instance, consisted of approximately 4700 MIDI notes. Once the MIDI clips are in *Live*, the *Live* set can then be rendered offline.

8 Conclusion and Future Plans

In this commentary I presented and discussed my audiovisual aesthetics, compositional and analytical models and my compositional workflow as a dialogue between ideas and technology.

I explored the juxtaposition of artificial and organic qualities on various levels of composition, ranging from basic compositional elements to the overall characteristics of the form. This was achieved by combining generative and conventional compositional approaches. Physics-based and pseudo physics-based algorithms were used to either deform pre-composed audiovisual structures or to generate audiovisual material that served as a starting point for further compositional decisions. Alternatively, generative material was forced onto the temporal and harmonic grid. A common denominator to all these scenarios was mutual adaptation of both generative and pre-composed elements from which interesting aesthetics emerged: audio-visual-temporal objects that slowly decay over time, awkward groove derived from the visual motion, sounds and rhythms reflecting physical qualities of geometric motion, time-stretched organic motion or explorations of continuous and fragmented time.

The idea of musical tension and release was cross-modally extended by the opposition of micro-isomorphic and macro-isomorphic relationships, using micro-isomorphic relationships to emphasize musical tension and macro-isomorphic relationships to emphasize release. In addition to that, the idea of *audiovisual escapology* was introduced that considers micro-isomorphic relationships as a cross-modal constraint. Contrasting predominantly micro-isomorphic relationships with macro-isomorphic ones results in the feeling of entrapment typical for gestalts that I call *monoids* while my concept of *multiplexes* offeres the sense of liberation in the form of detachment of sound and image from the object.

The aesthetic implications of different audiovisual hierarchies were explored and used in the context of *audiovisual escapology*. An approach in which geometry leads and audio follows was predominantly used to create tense audiovisual material characterized by tight audiovisual bonds and inherent musical tension. An approach where audio leads and geometry follows was used to either emphasize musical tension with micro-isomorphic gestures or to support musical release by matching the macro-qualities of music. Constant switching between audio and visual leads resulted in tight audiovisual material characterised by a fragmented musical language and contrasting visual gestures that were used to create audiovisual tension.

My audiovisual aesthetics have been well accepted in the world of intermedia art. The solo and collaborative works contained in portfolio received three international awards (*Edigma Semibreve Award* for *Spaceless Latitudes*, *Lumen Prize Student Award* for *Capillaries Capillaries*, *MADATAC Award for Most Promising Video Artist* for first nine minutes of *Capillaries Capillaries* which at the time existed as a self-standing piece called *Capillaries*) and were shortlisted for *Lumen Prize* in category moving image (*Spaceless Latitudes*). My first audiovisual performance was at the festival *Electric Spring* in Huddersfield in February 2017 and since then I have performed and exhibited at festivals such as *Ars Electronica*, *L.E.V. Festival*, *Brighton Digital Festival*, *Semibreve Festival*, *Athens Digital Arts Festival*, *Node Festival* or *Madatac*.

The plan for the future is to continue with the exploration of audiovisual aesthetics within the isomorphic audiovisual paradigm. I intend to extend my concept of audiovisual gestalts within different formats and mediums. The first step in this direction was already taken with the installation *Synspecies - Spaceless Latitudes* that exploits three video channels and 4.1 surround sound. The plan for December 2018 is a collaboration with *Ars Electronica*'s *FutureLab* on a project *Immersify*. As part of this project I will have the opportunity to explore stereoscopy for bigger displays from different perspectives – the idea is to make an active stereo 3D version of *Singing Sand* in 8K resolution for wall and floor projection in *Ars Electronica*'s *Deep Space 8K*. With Elías Merino, my collaborator on the project *Synspecies*, we are currently preparing the audiovisual material for the premiere of the performance that will take place at *Festival L.E.V.* 2019. The performance will include the use of fog and lights with the aim to expand the visual part of audiovisual gestalts into space. A similar attempt, with the additional use of lasers controlled by audio signals, will be the main challenge of the project planned for the second half of 2019. This will be a collaboration with an audiovisual artist Alberto Novello.

9 Bibliography

References

Adkins, M., Scott, R. and Tremblay, P. A. (2016) ‘Post-Acousmatic Practice: Re-evaluating Schaeffer’s heritage’, in *Organised Sound*. Cambridge University Press, 21(2), pp. 106–116.

Arnheim, R. (1986). *New essays on the psychology of art*. Univ. of California Press.

Asch, S.E. (1968): *Gestalt theory*. In D.L. Sills (Ed.), A. Johnson (Hon. Ed), Vol. 6, pp. 158-175.

Audissino, E. (2017). *Film / Music Analysis: A Film Studies Approach.* Cham: Springer Nature.

Auslander, P. (2013). ‘Sound and Vision: The Audio/Visual Economy of Musical Performance’, in Richardson, J., Gorbman, C. and Vernallis, C. eds., (2013). *The Oxford handbook of new audiovisual aesthetics*. Oxford University Press, pp. 605-622.

Birtwistle, A. (2017). Cinesonica: sounding film and video.

Buhler, J. Flinn, C. and Neumeyer, D., (2000). *Music and cinema*. Wesleyan University Press.

Burt, G. (1994). *The Art of Film Music: Special Emphasis on Hugo Friedhofer, Alex North, David Raksin, Leonard Rosenman*. UPNE

Cascone, K. (2002). ‘The Aesthetics of Failure: 'Post-Digital' Tendencies in Contemporary Computer Music’, *Computer Music Journal* 24(4), pp. 12-18.

Chen, Y.-C., Huang, P.-C., Woods, A. & Spence, C. (2016) When “Bouba” equals “Kiki”: Cultural commonalities and cultural differences in sound-shape correspondences [Online], in *Scientific Reports*, 6. Available: https://www.nature.com/articles/srep26681 [Accessed 20 June 2018].

Chen, Y.-C., & Spence, C. (2017) Assessing the Role of the 'Unity Assumption' on Multisensory Integration: A Review, in *Frontiers in Psychology*, 8(445), pp. 1-22. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5374162/ [Accessed 20 June 2018].

Coulter, J. (2010). Electroacoustic Music with Moving Images: the art of media pairing. *Organised Sound*, *15*(1), pp.26-34.

Chion, M. (1994). *Audiovision: Sound on Screen*. New York: Columbia University Press

Cook, N. (1998). *Analysing musical multimedia*. Oxford: Oxford University Press 2004.

Grierson, M.S. (2005). *Audiovisual composition* (Doctoral dissertation, University of Kent at Canterbury).

Ingalls, M. (2014). Soundflower (Version 2.0b2) [Computer program]. Available: https://github.com/mattingalls/Soundflower (Accessed 3 September 2012).

Kerins, M. (2010). *Beyond Dolby (stereo): cinema in the digital sound age*. Indiana University Press.

Langkjaer, B. (2015). 'Audiovisual Styling and the Film Experience: Prospects for Textual Analysis and Experimental Approaches to Understant the Perception of Sound and Music in Movies', in *Music and the Moving Image*. University of Illinois Press, 2(8), pp. 35–47.

Leap Motion. (2018). Developer Archive. [Online]. [10 October 2018]. Available: https://developer-archive.leapmotion.com/gallery/grab-strength

Lehar, S. (2008). The constructive aspect of visual perception.

Lipscomb, S.D. (1999). Cross-modal integration: Synchronization of auditory and visual components in simple and complex media. *The Journal of the Acoustical Society of America*, *105*(2), p.1274.

Luchins, A.S. and Luchins, E.H., (1999). Isomorphism in Gestalt theory: Comparison of Wertheimer's and Kohler's concepts. *Gestalt Theory*, *21*(3), pp.208-234.

Luten, E. (n.d.). *OpenGLBook.com,* online book. Available: http://openglbook.com/chapter-3-index-buffer-objects-and-primitive-types.html (Accessed: 19 November 2018)

Max Forum. (2016). Cycling '74. [User Forum] Available: https://cycling74.com/forums/share-eliminate-redundant-replicated-vertices-with-jit-gl-model-gl-multiple (Accessed 2 November 2018).

Monaco, J. and Lindroth, D. (2000). *How to read a film: the world of movies, media, and multimedia: language, history, theory*. Oxford University Press, USA.

Palmer, S. (1999). Color, Consciousness, and the Isomorphism Constraint. *Behavioral and Brain Sciences 22* (6):923-943.

Rogue Amoeba (2016) Loopback (Version 5.8.2) [Computer program]. Available: https://rogueamoeba.com/legacy (Accessed 3 September 2012).

Schreuder, D.A. (2014). *Vision and visual perception*. Archway publishing.

Slutsky, D.A. & Recanzone, G.H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), pp. 7-10.

Spence, C. (2015). Cross-Modal Perceptual Organization, in Wagemans, J. (eds.) *The Oxford Handbook of Perceptual Organization.* Oxford: Oxford Press, pp. 639-654.

Spilotros, L. & Parzy, E. (2010). *Semantic Congruence and Perception: An Experimental Investigation*. Copenhagen: Aalborg University Copenhagen. Available: https://projekter.aau.dk/projekter/files/32015327/REPORT%20FINAL%20RELEASE.pdf [Accessed 22 October 2018].

Spence, C., Sanabria, D., and Soto-Faraco, S. (2007). 'Intersensory Gestalten and crossmodal scene perception', in K. Noguchi (eds.), *Psychology of beauty and Kansei: New horizons of Gestalt perception*. Tokyo: Fuzanbo International, pp. 519-57

Tsilionis, E. & Vatakis, A. (2016). Multisensory binding: is the contribution of synchrony and semantic congruency obligatory?. *Current Opinion in Behavioral Sciences*, *8*, pp.7-13.

Talsma, D., Doty, T.J. and Woldorff, M.G. (2006). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cerebral cortex*, *17*(3), pp.679-690.

Art Works Cited

Borowski, J. (2015). *Lissajous #2* [Audiovisual Composition]. Available: https://vimeo.com/125565277 (Accessed: 19 November 2018).

Cooper, M. & Lomas, A. (2014). *Seething / Cellular Forms* [Music Video]. Available: https://vimeo.com/98435385 (Accessed: 19 November 2018).

Coulter, J. (2009). *Abide With Me* [Audiovisual Composition]. Available: https://vimeo.com/49351400 (Accessed: 19 November 2018).

DePalma, B. (1976). *Obsession* [Movie].

Droljc, T. (2014). *Particle-based Granular Synthesis 1* [Video]. Available: https://www.youtube.com/watch?v=b5xZ5Urc3kY (Accessed: 19 November 2018).

Droljc, T. (2015). *Stringvencer 1.0* [Video]. Available: https://www.youtube.com/watch?v=JpdGQbkCV_o (Accessed: 19 November 2018).

Droljc, T. (2017). *Particle-based Granular Synthesis 2* [Video]. Available: https://www.youtube.com/watch?v=ajCYefU61e8 (Accessed: 19 November 2018).

Frost, B. & Weber, M. (2014). *Aurora* [Audiovisual Composition]. Available: https://vimeo.com/111845602 (Accessed: 19 November 2018).

Goffey, N. & Hawley, D. (2005). *Chemical Brothers - Believe* [Music Video].

Jebanasam, P. & Barri, T. (2016). *Continuum AV (fragment)* [Audiovisual Composition]. Available: https://vimeo.com/196428018 (Accessed: 19 November 2018).

Kolgen, H. (2014). *Seismik – Perfo Extracts* [Audiovisual Performance]. Available: https://vimeo.com/102158828 (Accessed: 19 November 2018).

Kurokawa, R. (2014). *Ground.alt* [Audiovisual Composition]. Available: https://vimeo.com/93334650 (Accessed: 19 November 2018).

Kurokawa, R. (2014). *Syn_mod.2* [Audiovisual Composition]. Available: https://www.seditionart.com/ryoichi-kurokawa/syn_mod-2 (Accessed: 19 November 2018).

Kurokawa, R. (2016). *Unfold* [Audiovisual Composition]. Available: https://vimeo.com/160648305 (Accessed: 19 November 2018).

Moura, J. M. (2011). *SuperCollider Shape* [Audiovisual Composition]. Available: https://vimeo.com/28275072 (Accessed: 19 November 2018).

Shadforth, D. (1999). *Moloko – Sing it Back* [Music Video].

Spielberg, S. (1991). *Hook* [Movie].

Stern, W. (1996). *Prodigy - Firestarter* [Music Video].

Tanaka, A. (2008). *New Biomuse Demo* [Video]. Available: https://vimeo.com/2483259 (Accessed: 19 November 2018).