Audiovisual Particles

Parameter Mapping as a Framework for Audiovisual Composition

STEPHEN CALLEAR

A thesis submitted in partial fulfilment of the
requirements of Bath Spa University for the
degree of Doctor of Philosophy

School of Music and Performing Arts, Bath Spa University

November 2012

Abstract

This thesis investigates the role of cross-modal correspondence within audiovisual composition, presenting both a conceptual model and a methodological framework for the creation of abstract audiovisual art. While this research is specifically aimed at the field of abstract digital animation it is also intended to act as a platform for the future development of concurrent audiovisual synthesis techniques within the general field of audiovisual art.

Referencing literature regarding the psychophysiological bases for audiovisual integration, it is argued that temporal congruence offers a mechanism for the manipulation of cross-modal correspondence within audiovisual media. Further to this, electroacoustic and formalist theory is discussed with specific reference to the interrelationship of medium structures to enable the identification of a conceptual model for audiovisual composition.

Referencing theory from the fields of musical instrument design and algorithmic composition, parameter mapping is identified as a mechanism for the modulation of temporal congruence. Its implementation within audiovisual composition is then discussed. Derived from both this and a conceptual parallel between the organisational structures of audio grains and visual particles, the *audiovisual particles* framework is presented as a methodological basis for the creation of abstract audiovisual art.

The presented theory is supported by a series of demonstrative studies exploring both the practical application of the audiovisual particles framework and the role of parameter mapping within the process of audiovisual media generation. Experiential observations are discussed for each to inform future praxis. In addition, two audiovisual compositions are presented as both implementations of developed theory and as artworks in their own right.

Acknowledgements

Foremost, I would like to express my sincere thanks to my supervisor Professor Joseph Hyde, whose encouragement, guidance and support has proved essential from the beginning to the conclusion of this research. I also wish to thank my second supervisor Dr Andy Keep, whose advice and testing questions have been consistently invaluable, especially during the crucial formative stages of the research process. I would also like to extend my thanks to all staff within the Bath Spa University School of Music and Performing Arts who have offered me advice, assistance or opportunity throughout the duration of this research.

I would like to express my gratitude to Verity Stoffell for both her support and advice and also her assistance in gathering much of the source audio material that has proved so essential to the practical component of this work. Finally, I would like to thank my parents for their support, advice and encouragement, without which none of this would have been possible.

Index

---

1 Introduction

1.1 STATEMENT OF OBJECTIVES

The integration of sound and image within audiovisual media has been the focus of sustained theoretical development throughout the twentieth century. Explicit discussion of compositional approaches for the creation of such art, however, remains relatively rare. Indeed, whilst the field of abstract audiovisual art continues to thrive at the forefront of technological development and there exists a healthy, if divergent, literature basis for the discussion of audiovisual correspondence, there has been minimal development regarding the synthesis of such discourse into either a conceptual model or methodological framework. On this basis, the primary aims of this research are as follows:

- The exploration of cross-modal correspondence within the context of audiovisual art to permit the identification of a theoretical basis for its manipulation.
- The application of parameter mapping theory from the fields of digital instrument design and algorithmic composition to the process of audiovisual media generation.
- The development of a methodological framework for the creation of audiovisual art.
- The practical application of proposed theory within a series of audiovisual studies and artworks.

The intention of this research is to contribute to the field of abstract digital animation, in which there is a relative dearth of literature regarding methodological approaches to the integration of sound and image. Further to this, it is hoped that this thesis may contribute to the future development of concurrent audiovisual synthesis techniques and multi-scale composition methodologies within the more general field of audiovisual art.

1.2 OVERVIEW

The thesis begins with a definition of the contextual basis for the presented theory and praxis. To this end, the art and theory of animators John and James Whitney are discussed, with a specific focus on the theories of audiovisual harmony presented by the former. Following this, the works of a selection of abstract audiovisual artists are critiqued to contextualise the theoretical and aesthetic motivation behind this research.

In Chapter 3, theory regarding audiovisual correspondence within the context of perception and cognition is discussed. Within this, spatial, temporal and semantic modes of integration are presented and the relevance of each is assessed in the context of this research. Developing from this, semantic and temporal congruence are identified as measures of cross-modal integration within the context of audiovisual art.

Chapter 4 begins with an overview of the potential application of electroacoustic theory to the field of visual composition. This is followed by a review of theory regarding temporal structure in the context of electroacoustic composition; this is then used to inform a discussion of temporal morphology within audiovisual composition. Prevalent formalist theory is discussed next, with specific reference to temporal congruence to enable a multi-scale model for audiovisual composition to be developed.

In Chapter 5, the role of parameter mapping within the creation of audiovisual media is examined. The chapter begins with the identification of parameter mapping terminology prevalent within the fields of instrument design and algorithmic composition. Following this the concept of mapping hierarchy is presented and theory regarding the perceptual transparency of mappings is discussed in both general terms and in the context of audiovisual composition. This is followed by a discussion of techniques for the modulation of mapping transparency within audiovisual composition. Finally, an overview of media-driven and data-driven mapping hierarchies is presented as precursor to Chapter 6.

Derived from theory presented in Chapters 4 and 5, Chapter 6 presents the *audiovisual particles* framework. The aim of the framework is to define a methodological approach for the creation and organisation of audiovisual gesture which will allow the formation and dissolution of audiovisual congruence within audiovisual art. The chapter begins by drawing a conceptual parallel between audio grains and visual particles, followed by the identification of concurrent synthesis techniques as a potential mechanism for the creation of temporal congruence. The concept of density is then presented as a method for analysing particle structures. This is followed by the identification of data-driven composition techniques as a mechanism for the transposition of morphological processes from data to synthesis systems. Finally the component processes of the audiovisual particles framework – material selection, media generation and media organisation – are discussed in detail.

In Chapter 7 a series of composition studies are presented as example implementations of the audiovisual particles framework. The composition methodology for each study is assessed to allow the practical application of the presented framework to be discussed. Drawing from this analysis, the relevance of implicit and explicit mapping strategy within the media generation process is evaluated and observations applicable to the process of media organisation are discussed.

As an extension to the compositions presented in Chapter 7, Chapter 8 documents four additional studies which analyse parameter mapping strategy within the process of audiovisual media generation. The methodology for each study is presented alongside observations made throughout their creation. This is followed by cross-analysis of the resultant works and the adopted mapping strategy is discussed within the context of audiovisual art. A distinction is then drawn between explicit and implicit particle relationships and the relative merits of each approach are evaluated.

In Chapter 9, the works *In Perpetuity: The Early Lilacs* and *In Perpetuity: The Linden Trees* are presented as practical applications of theory developed throughout this research process. The chapter begins with an overview of the software tools that have been developed and the definition of a generic composition methodology. The adopted media generation processes are then documented, detailing the particle simulation and parameter mapping systems used throughout the creation of each composition. This is followed by a detailed discussion of the final artworks with specific reference to the methodological, conceptual and aesthetic aims of this research.

The thesis concludes with an evaluation of presented theory and praxis with direct reference to established research aims. Further to this, potential directions for the future development and application of the proposed theory are discussed.

2 Audiovisual Context

2.1 INTRODUCTION

This chapter discusses the works and theory of a selection of abstract audiovisual composers as a means of contextualising the theoretical and aesthetic bases for this research. The chapter begins with an overview of the work and theory of animators John and James Whitney, with a specific focus on the theories of audiovisual harmony presented by the former. Following this the compositions of contemporary audiovisual artists Jean Piché, Bret Battey and Dennis Miller are critiqued to establish the context of this research within the field of abstract digital animation. The chapter concludes with a discussion of pertinent theoretical and aesthetic concepts that may inform the development of theory and composition methodology.

2.2 DIGITAL HARMONY

The mapping of sound and image has been a topic of sustained research throughout the twentieth century (Brougher, 2005; Strick, 2005; Zilczer, 1987; 2005). Ranging from the painted works of Kandinsky and Klee to the abstract animation of Fischinger and McLaren, the field of visual-music has been a focal point for the discussion of cross-modal correspondence. Within the context of this research, the art and theory of John and James Whitney throughout the latter half of the twentieth century represents the most significant development in audiovisual composition methodology and offers a strong basis for the contextualisation of this research.

With their *Five Film Exercises* (1943-44), brothers John and James Whitney coupled pendulum controlled audio synthesis with stencilled light to create an 'almost ruthless … assault on eye and ear' in which 'the images and sound seem inextricably linked' (Brougher, 2005:125). Marking a distinct movement away from 'the overly familiar forms derived from Kandinsky and reprised by Fischinger' (*ibid.*), James Whitney's *Yantra* (1950-57) employs particles of light, multiplied and grouped to create complex evolving structures. With his later work *Lapis* (1963-66), the 'complex macrocosmic forms' that result from this technique create what Brougher (2005:132) describes as 'one of the most convincing works in the history of modern art.' Informed by twelve-tone serial composition techniques (Garro, 2005) and disregarding the direct synaesthetic mappings prevalent within many visual music works of the early twentieth century (Brougher, 2005), John Whitney's compositions explore contrapuntal relationships between medium events. This is perhaps most explicitly demonstrated in his works *Catalog* (1961) and *Permutations* (1968).

This rejection of the 'illustrative' function of video within audiovisual composition forms the basis for John Whitney's theories of cross-modal harmony (Whitney, 1980; 1991; 1994). In his text *Digital Harmony: On the Complementarity of Music and Visual Art* (1980:26), Whitney notes a distinction between the compositional processes of aural and visual material, observing that 'no visual motion [works] the way musical motion works.' Instead, he hypothesises that 'the relationships of sight and sound would be best served if it were possible to compose both components … within some common aesthetic such as harmony would offer' (*ibid.*). This theory of a 'differential digital harmony' is formed on the basis that 'progressions of ratio in visual as well as tonal (chordal) patterns of harmony lie at the heart of our perception of time as aesthetic structure' (Whitney, 1994:48). The associative relationship that results from this proportional symmetry of structures is termed by Whitney as a 'complementarity' (Whitney, 1980; 1994), permitting what Alves (2005:46) describes as a 'multidimensional interplay of tension and resolution.'

The importance of computer algorithms in the creation of ratiometric audiovisual structures is discussed at length in Whitney's essay *To Paint on Water: The Audiovisual Duet of Complementarity* (1994). Of particular relevance to this research is Whitney's argument that a ratiometric complementarity may be formed between the products of granular synthesis and visual particle systems due to structural similarities between each. Whitney proposes that this equivalence may enable the creation of cross-modal counterpoint and a 'richer virtual reality' (Whitney, 1994:51) within audiovisual art.

2.3 CONTEMPORARY AUDIOVISUAL ART

The field of contemporary audiovisual art spans a diverse range of genres including, but not limited to, music video (Fahlenbrach, 2005), live cinema/VJ performance (Faulkner, 2006; Alexander and Collins, 2007; Jones, 2008) and interactive/installation works (Sommerer and Mignonneau, 1999; Edmonds *et al.*, 2004; Edmonds and Pauletto, 2004; Oliver and Pickles, 2007). While elements of this research rely heavily upon literature from these fields, of greater importance to the contextualisation of the proposed theory is the field of abstract digital animation. Within this categorisation notable artists include Jean Piché, Bret Battey and Dennis Miller. While this list is far from exhaustive, works by these composers exemplify the fundamental methodological and aesthetic bases for the proposed research: the composer-led generation and integration of abstract visual form and sound to create a unified audiovisual output.

2.3.1 Jean Piché

An established electroacoustic composer, Jean Piché's video works are the product of a non-narrative audiovisual form described by the artist as 'vidéo-musique' (Piché, 2004; Steenhuisen, 2009). Transposing techniques developed from composing with sonic material, Piché's works often employ processed *concrète* material with limited referential context. This approach to composition is perhaps best demonstrated by *Sieves* (2004), in which the artist combines manipulated audio samples and granular textures with images of the ground showing 'intricate details of the soil' (Steenhuisen, 2009:262). The first half of the work exhibits gentle modulations in sonic timbre that complement the evolutions in visual form and texture. The use of emergency transmission recordings in the latter half of the work adds 'dramatic texture' (*ibid.*) that correlates with the rapid changes in visual form and colour. Strong synchronisation between aural and visual components is maintained throughout this section, creating a rhythmic sense of pace that reinforces the perceived urgency of the audio material.

Continuing the electroacoustic idiom established in *Sieves*, Piché's installation works *Boréales* (2008 – 2009) and *Australes* (2010 – 2011) employ abstract visuals generated by using recorded video imagery to control the parameters of a particle generator. As a result, the visuals exhibit slow transformations in form and texture, morphing from states of abstract texture to apparent mass. Interestingly, while *Boréales* demonstrates a relatively limited visual palette in comparison to *Australes*, the audio component possesses greater spectral diversity, morphing between subtle rhythms and ethereal textures, interspersed with processed vocal samples. Conversely, the faster modulations in visual form exhibited throughout *Australes* are accompanied by a relatively sparse spectral palette with gentler modulations in timbre. In comparison to *Sieves*, both works exhibit a more dynamic temporal relationship between sound and video, with the medium structures converging upon points of synchronisation before diverging into independent morphologies.

Piché's most recent work, *Océanes* (2010 – 2011), moves away from the methodology established in previous compositions, employing synthesised audio instead of processed recordings. Generated with formant synthesis, the audio component of the work exhibits subtle modulations in spectral content, giving it an ethereal tone that matches the slow evolutions of form present within the video. Analogous to the granular nature of the audio component (an observation made by the composer in the notes for the piece), the visual medium is generated using particle synthesis, creating textural waves that correspond to the title of the work. Impressively, while the fundamental visual structures remain relatively consistent throughout the duration of the work, the textural evolutions of the imagery create a strong sense of tonal depth and variation. Similarly, modulations in visual density create complex variations in the visual form, often directly complementing the spectral structure of the audio component. As with *Boréales* and *Australes*, *Océanes* exhibits only limited elements of synchronisation, instead presenting more gestural correspondences between medium morphologies.

2.3.2 Bret Battey

The audiovisual work of Bret Battey combines electroacoustic composition techniques with generative functions to create abstract audiovisual form. For example, in *Autarkeia Aggregatum* (2005) the visual component is created using a custom visual filter that regenerates and modulates the pixel data of source imagery to create new forms (Battey, 2006). This use of source imagery is similar to the methodology employed in many of Piché's audiovisual works, with the processing of recorded video being analogous to the processing of audio material within many electroacoustic works. The fluid transformations in visual form create a sense of constant motion, with the visual texture gently morphing between states. The visual filter contributes strongly towards this, with the component particles appearing as autonomous entities at times and as a single texture at others. The soundtrack complements the gentle morphology of the visuals, exhibiting subtle transformations in spectral content that often directly correspond their temporal morphology.

Battey's *Luna* series of works comprises three compositions: *Mercurius* (2007), *Lacus Temporis* (2008) and *Sinus Aestum* (2009). Adopting the same video filtering process, *Mercurius* presents a similar visual aesthetic to *Autarkeia Aggregatum*, although the inclusion of additional particle behaviours creates a wider palette of visual forms. The audio component of the piece exhibits slow transformations from spectrally simple tones to dense timbres and noise components. While the temporal morphology of the audio complements the transformations in visual form, explicit synchronisation of medium structures is rare. Indeed, relative to *Autarkeia Aggregatum* the composition incorporates less explicit correspondences between the spectral and visual complexity. *Lacus Temporis* (2008) exhibits a more minimalist aesthetic relative to *Mercurius*, with subtle changes in visual form and a low-contrast colour palette. The audio component of the work employs feedback synthesis to create a transforming, ethereal sonic texture that modulates between noise and pitched components. In comparison to preceding works, the audiovisual relationships in *Lacus Temporis* often appear to be more literal, with changes in visual density often being accompanied with an equivalent modulation in the spectral density of the soundtrack. Further to this, many visual transformations are complemented by synchronised modulations in audio pitch and timbre. *Sinus Aestum* (2009) is perhaps the most complex of the *Luna* series. The visuals transform from dense textural forms to complex spiral structures giving the impression of increased depth relative to both *Mercurius* and *Lacus Temporis*. The audio component complements these transformations, morphing between pitched tones and spectrally dense textures in a manner that subtly corresponds to the visual transformations without directly replicating them.

2.3.3 Dennis Miller

Audiovisual works by composer and animator Dennis Miller typically combine abstract digital imagery with spectrally complex sonic material. *White Noise* (2007) incorporates three-dimensional visual elements with an audio component that oscillates between tonal pitches, spectrally complex sonic textures and bursts of noise. Complementing this, the visual perspective shifts abruptly between textural forms and geometric entities, themselves the product of complex and rapid transformations in form and colour. In general, the cross-modal correspondences are subtle, allowing the morphologies of each medium to complement the other without explicit synchronisation. In its most dramatic gestures however, transformations in audio amplitude and spectral density are replicated in the visual morphology, contributing to the impact of the piece.

In comparison to *White Noise* the visual component of *Lines of Force* (2008) is more textural, with limited apparent dimensionality. The visual component evolves in a fluid manner, but with a measured pace that complements the strong rhythmic elements within the audio. The colour palette is comprised primarily of browns and blues that correspond thematically to the 'industrial' aesthetic implied by the timbre and rhythms of the audio component. While there are elements of temporal synchronisation in many of the recurring gestures, the piece incorporates strong divergences between medium structures, with the latter half of the work often combining dense noise-based audio with a monochrome colour palette and limited visual morphology. There are occasional spatial correspondences, with visual motion corresponding to a perceived spectral or pitch trajectory within the audio.

Although structurally quite different, *Echoing Spaces* (2009) employs a similar visual aesthetic to *Lines of Force*, but with limited rhythmic elements and slower transformations in form. The colour palette remains relatively constant throughout the work, with only the saturation and brightness of the image changing significantly. This results in a calmer aesthetic than *White Noise*, reinforced by more subtle transformations in both spectral complexity and pitch. While the piece exhibits occasional correspondences between the temporal structures of each medium, the morphologies appear to develop independently for much of the work, converging only briefly. Instead, the piece exhibits a more general correlation between medium densities, with each sonic gesture typically being associated with a different visual form.

2.4 DISCUSSION

The art and theory of John and James Whitney forms a strong contextual basis for this research. Most significant are the theories of harmony, counterpoint and complementarity presented in John Whitney's written work (1980; 1991; 1994), indicating the potential for complex cross-modal relationships derived from 'contextualised rules' (Garro, 2005:3). Whitney's discussion of computer generated media and algorithmic composition within his latter texts (1991; 1994) also indicates the potential for digital animation techniques in the creation of abstract art. The theoretical parallel Whitney (1994) draws between granular synthesis and visual particle systems is also significant within the context of this research, indicating the potential for a unified methodological process that encompasses both systems. This concept forms the basis for both the theoretical and practical components of this research.

In each of the contemporary audiovisual works discussed above, the artists compose both aural and visual components, forming cross-modal interactions between morphologies established in each medium. Although only explicitly discussed by Piché (Steenhuisen, 2009), the translation of electroacoustic composition methodology to the creation of the visual component is prevalent within the discussed works. Of particular note here is the processing of referential imagery to create abstract form in both Piché and Battey's compositions, a methodology that this research considers to be analogous to the sonic reductions of Schaeffer's *musique concrète* (see Chion, 1983; Kane, 2007). Similarly, a parallel can be established between the textural transformations in visual form present in the discussed works and the morphology of sonic properties within electroacoustic composition. For example, the modulations in visual density throughout *Sinus Aestum* (Battey, 2009) could be applied to the processing of spectral density within sonic composition and vice versa, a technique that Battey adopts with great success throughout the work. This thesis is both informed by and expands upon this methodological precedent, drawing from the field of electroacoustic composition to establish a composition framework for the creation of abstract audiovisual art.

The works discussed adopt a relatively consistent audiovisual aesthetic that informs both the stylistic and methodological intentions of this research. The use of particle systems in the generation of the visual component by both Piché and Battey is fundamental to the creation of non-referential, textural forms that imply mass and dimensionality through complex evolutions in particle states. Similarly, whilst implying greater structural 'solidity', the three-dimensional forms within Miller's *White Noise* (2007) evolve in a manner that encourages the evaluation of visual properties rather than any referential basis. In general, the discussed works make limited use of perspective or camera movements in the creation of visual depth. Instead, this is typically implied through textural complexity. For example, the modulations in visual detail in Piché's *Oceanes* (2010-2011) imply a multi-layered sense of depth within the fundamentally two-dimensional imagery. Similar effects can also be seen throughout much of Battey's work.

All of the compositions incorporate primarily textural audio components that evolve between states of spectral simplicity and complexity. As with the visuals, the audio components remain fundamentally non-referential, the only exception being Piché's use of processed vocal recordings throughout many of his works.

With *White Noise* (Miller, 2007) being a notable exception, the discussed works generally exhibit slow transformations in audiovisual form, incorporating evolving aural and visual components with gestural accents. The structures of each medium typically develop independently, converging and diverging to suit the overall structure of the work. There is however, some use of explicit synchronisation, often providing greater impact to a particular audiovisual gesture. For example, within Miller's *White Noise* (2007), sections of greater spectral density and impact are typically more strongly synchronised with the visual component than in other sections of the work. This correspondence between aural and visual densities is also apparent across the other composers' work in less explicitly synchronised contexts. Indeed, all of the works incorporate gestural correspondences between the spectral complexity of the audio and the visual activity or density. This complex, dynamic relationship between medium structures conforms to Whitney's theory of 'complementarity' discussed above. As a result, the relationship between medium structures forms a core component of this research and is the subject of continued discussion throughout this thesis.

Despite significant discussion of the subject within twentieth century arts literature (Garner, 1978; Davis, 1979; Wells, 1980; DeWitt, 1987; Peacock, 1988; Alves, 2005), there appears to be no direct correlation between visual colour and audio spectrum or pitch in any of the works. Instead, a more thematic correspondence is apparent. For example, noise-based textures are combined with an intense, desaturated colour palette in *White Noise* (Miller, 2007), while in *Autarkeia Aggregatum* (Battey, 2005) pitched, ethereal tones are coupled with soft blue hues. In the context of the discussed works, this research therefore considers that the thematic use of colour is more meaningful than any explicit mapping between audio properties and colour hue, a position that is maintained within the composition framework and practical components of this thesis.

2.5 CONCLUSION

In conclusion, an overview of the art and theory of John and James Whitney has been provided to contextualise the contemporary field of abstract digital animation. Within this field, the artists Jean Piché, Bret Battey and Dennis Miller have been identified as key proponents of the aesthetic and methodological processes relevant to this research. Of particular note is the transposition of electroacoustic composition techniques to the visual medium by both Piché and Battey. This topic will be the subject of further discussion in Chapter 4. Further to this, the audiovisual aesthetic adopted by each of the artists has been identified as a basis for the development of composition methodology within this research.

3 Perception and Cognition

3.1 INTRODUCTION

Referencing psychological and neurological theory, this chapter explores audiovisual perception and cognition as a mechanism to inform an objective and interpretable correspondence between auditory and visual events. The chapter begins with an overview of the cognitive interaction between auditory and visual stimuli, with particular reference to the phenomena of spatial ventriloquism, temporal ventriloquism and cross-modal interference. This is followed by a discussion of theory regarding audiovisual congruence. Within this, mechanisms for the formation of both temporal and semantic correspondences are investigated and the potential implications of such theory within the perception of audiovisual media are discussed. The chapter concludes with an overview of presented theory and the identification of phenomena relevant to the composition and perception of audiovisual art.

3.2 CROSS-MODAL INTERACTION

Theory regarding the interaction between visual and auditory modalities is well established (Spence, 2007). From a young age the human brain responds to temporal correlations in vision and audition using ‘each sense to correct the scene-analysis decisions of the other’ (Bregman, 1994:181). Spatial ventriloquism is a phenomenon in which spatially disparate auditory and visual stimuli are perceived to occur in the same location when presented in synchrony (Welch and Warren, 1980), with the strength of the effect decreasing as spatial disparity or temporal asynchrony is increased (Spence, 2007). Typically vision forms the dominant stimulus with auditory events being localised towards the visual event, although effects in both directions have been observed (Welch and Warren, 1980; Bregman, 1994).

Similarly, temporal ventriloquism is a phenomenon in which ‘visual stimuli appear to be “pulled” into approximate temporal alignment with corresponding auditory stimuli’ (Spence, 2007:61) when presented within a certain temporal window (Lewkowicz, 1999; Spence and Squire, 2003). Research suggests that both subject-stimuli distance and subject adaptation have an effect upon the temporal limits of perceived audiovisual synchrony (Sugita and Suzuki, 2003; Fujisaka *et al.*, 2004). Cross-modal interference is also evidenced by the ‘McGurk effect.’ When a videotape of a face pronouncing a consonant-vowel syllable such as ‘ga-ga’ is played in synchrony with a soundtrack containing altered phonemes such as ‘ba-ba’, subjects generally perceive an illusory blend of the two senses, in this case ‘da-da’ (Bregman, 1994).

This research considers that three generalised observations relevant to the field of audiovisual art may be derived from the presented phenomena. Firstly, spatial ventriloquism indicates that temporal synchronisation takes precedence over spatial synchronisation, thus permitting the formation of cross-modal associations between visual and aural events presented in synchrony despite significant disparity between their locations. Secondly, theory regarding temporal ventriloquism indicates that a window of asynchrony is permitted before events in each medium appear exclusive. Lastly, the McGurk effect indicates that events in one modality may alter our perception of the other when presented in synchrony, although explicit investigation of this phenomenon is outside the scope of this research.

3.3 AUDIOVISUAL CONGRUENCE

On the basis that both auditory and visual mediums can be analysed in terms of structural features and perceived meaning (Cohen, 1998; 2001; 2005), this research posits that audiovisual correspondence can be explained in terms of perceptual correlations between these elements. As such, the following section addresses theory regarding temporal and semantic congruence in order to evaluate its relevance within the composition and perception of audiovisual media.

Cohen (1998:14) defines the temporal structures of both sound and image as the 'rhythm, contour or patterning in time' of the formal properties of each. The perceived alignment of structural features permits the identification of temporal congruence: the perception of correspondences between the temporal morphologies to facilitate cross-modal association. Lipscomb and Kendall (1994) observe that this mapping of 'periodicities' occurs throughout the process of perception in order to permit data reduction, while Cohen (1998; 2005) puts forward the importance of innate cognitive grouping principles (see Snyder, 2000) in the cross-analysis of temporal structure.

Semantic congruence is the evaluation of media 'appropriateness' based on semantic connotations. Citing Marshall and Cohen (1988), Lipscomb and Kendall (1994:90) refer to this phenomenon as the 'association judgement,' asserting that the perception of semantic congruence 'relies on past experience as a basis for determining whether or not the music is appropriate for a given context.' Cohen (1993:165) argues that 'the effects of combining music and film depend on the summation of … activated elementary percepts and emotions.' On this basis Cohen (*ibid.*) observes that the 'multimodal phenomena of film music can be understood in terms of connections formed between these elements.'

The effects of semantic and temporal congruence upon the perception of audiovisual media have been the subject of a wide range of empirical research. A study by Marshall and Cohen (1988) investigates the effects of musical soundtracks on the perceived characteristics of animated geometric forms. Within this experiment, subjects were randomly assigned one of a selection of music and film combinations and would rate them on a series of adjective rating scales. The results of the study indicate that the perceived meaning of the visual material is modified by the musical soundtrack through a process of association. Further to this, the authors postulate that temporal congruence between the structures of the visual and aural components contribute to the interpreted meaning and may alter the attentional strategy of the participant with regards to the audiovisual composite. The combined effect of temporal congruence and association of meaning forms the basis for the presentation of the Congruence-Associationist Framework for the analysis of the effect of soundtracks in audiovisual media. Further development of this framework is presented by Cohen (1998; 2001; 2005).

A later study by Cohen (1993) explores the strength of semantic congruence in isolation. In the experiment, a ball bouncing at a varying height and tempo was played alongside an audio track of varying pitch, tempo and melodic variation. Subjects where then asked to assess the visual component in terms of sadness/happiness. The results of the experiment indicate that the different soundtrack selections influenced the perceived rating of the visual entity. A similar study by Lipscomb and Kendall (1994) investigates the perceived relationship between musical soundtracks and visual images. In this, five visual scenes were selected from a commercial film along with the original score. Each soundtrack was then paired with each visual scene and participants were asked to select the 'best-fit' composite. Further to this, participants rated each of the scenes against a series of adjective scales. The majority of the subjects selected the original soundtrack for each scene, leading the authors to conclude that the judgement of the audiovisual composite is affected by the 'appropriateness' of the media pairing.

This concept is also examined by Bolivar *et al.* (1994). Within this study, footage of both friendly and aggressive wolf interactions was paired with friendly and aggressive musical soundtracks. Participants were then asked to rate the perceived semantic congruence for each of the pairings. Perhaps unsurprisingly, friendly-friendly and aggressive-aggressive audiovisual pairings where generally rated as congruent, while friendly-aggressive and aggressive-friendly pairings where rated as incongruent. Within the same study, an additional experiment asked participants to rate the perceived friendliness/aggressiveness of the wolf interactions when played with both semantically congruent and incongruent audio. The results of this experiment indicate that the soundtrack consistently modified the perceived mood of the visual component. For example, interaction between the wolves may be judged as aggressive when played alongside aggressive music, while the same behaviour may be deemed friendly if paired with friendly music. It is noted, however, that this is dependent on an element of ambiguity within the visual scene: 'it appears that when a scene provides only one interpretation, the meaning of the music is not needed or called upon for clarification' (Bolivar *et al.* 1994:48).

The influence of semantic congruence upon object recognition is explored in Molholm *et al.* (2004). Within this study, participants were played a range of congruent and incongruent animal picture and vocalisation pairings and the time taken for a particular visual object to be correctly identified was recorded. The results of this research indicate that congruent media pairings enable more rapid identification of a pre-specified object than pairings which are semantically incongruent. For example, the target stimuli of a cow was identified more rapidly and accurately when an image of a cow was presented alongside a mooing sound than when the same image was presented with the sound of a dog barking. In discussing this research, Spence (2007) notes that such observations do not necessarily imply that the aural and visual media are being integrated in any meaningful way. The theory does, however, reinforce the notion that when presented with stimuli in each medium, the brain refers to pre-existing knowledge to establish the perceived congruence of the pairing.

Semantic congruence is also seen to have an effect upon the memorability of audiovisual events. In a study by Boltz (2004), the visual memory and music recognition of subjects was tested in relation to a selection of semantically congruent and incongruent media pairings. The results of this research indicate that semantically incongruent pairings result in non-integrated memory representations, while semantically congruent pairings result in a unified memory representation. On this basis, the authors conclude that the memorability of audiovisual information is influenced by the relative appropriateness of media pairings. Interestingly, visuals consistently formed the dominant medium throughout the experiment. A later study by Boltz *et al.* (2009) expands on this, concluding that both semantic connotations and the temporal structure of the visual component affect the perception of the soundtrack in audiovisual pairings. It is noted, however, that while the perceived temporal structure of the audio was affected by the visuals, the pitch quality was not.

While the relevance of correlations between the temporal structures is regularly discussed in the above studies, none explicitly disregard semantic connotations formed through the use of referential imagery. In contrast, a study by Iwamiya and Ozaki (2004) explores the effect of temporal congruence between visual morphology and the pitch pattern of sound on the perception of the audiovisual composite. Within this, visual transformations of geometric shapes were played alongside audio components with a variety of pitch and amplitude patterns. The results of the study indicate that structural congruency can be achieved through ‘equivalent’ audiovisual transformation. Further to this, the authors observe that temporal congruence encourages increased evaluation of the visual component.

A consistent observation throughout the reviewed literature is that the structural and semantic characteristics of the auditory medium can influence the perceived structure and mood of the visual component (Marshall and Cohen, 1988; Bolivar *et al.*, 1994; Lipscomb and Kendall, 1994; Boltz, 2004). While the majority of studies address the effect of audio upon the visual perception, research by Boltz *et al.* (2009) indicates that the visual component is also affected by the perceived structure of the audio. This is unsurprising considering the theories regarding cross-modal interaction presented at the beginning of this chapter.

Similarly, research consistently indicates that events that are temporally congruent hold greater significance and are more rapidly identified than those that are incongruent (Marshall and Cohen, 1988; Lipscomb and Kendall, 1994; Cohen, 1998; Cohen 2001; Lipscomb and

Tolchinsky, 2004). This suggests that events that exhibit similar temporal accent structures within a complex audiovisual scene will be more rapidly and clearly associated than those with differing accent structures. Observations by Boltz (2004) indicate that the same effect occurs with semantic pairings. Extrapolating from this, this research considers that aural and visual media within a complex audiovisual scene will be more rapidly associated if derived from the same referential basis. Similarly, semantic congruence may also facilitate the manipulation of audience attention within an audiovisual work. For example, on the basis of research presented by Molholm *et al.* (2004), if a visual scene consists of images of a range of animals and the sound of one of the animals is also present, it is logical that attention will be directed towards the semantically congruent pairing. It should be noted however, that the significance of semantic congruence will be subjective: 'the decision will be determined by the subject's past experience and is, as a result, individual-specific' (Lipscomb and Kendall, 1994).

Also of interest is the consistent observation that the identification of congruence in one form may encourage the perception of congruence in another. For example, the majority of the studies observe that the identification of correspondences between temporal structures facilitates the perception of semantic congruence as attention is drawn to the congruent pairing (Marshall and Cohen, 1988; Lipscomb and Kendall, 1994; Cohen, 1998; Cohen 2001; Lipscomb and Tolchinsky, 2004). As Cohen (2005:26) notes, 'if there is no match between the audio and visual accent structure, then the direction of attention will shift from the audiovisual pair and associative information from the audio source will not be transmitted.' The reverse effect is observed by Bolivar *et al.* (1994:44), who note that 'the salience of the semantic congruency influenced judgements of temporal congruency.' Theory for this cognitive response is offered by Lipscomb and Kendall (1994:91) who observe that 'if the composite is determined inappropriate … the subject is more likely to separate the unit into its two constituent perceptual modalities in an explicit analytical attempt to determine a reason for the incongruence.'

On this basis, three primary observations can be derived from the presented studies. Firstly, both temporal accent structure and semantic connotations of one medium can influence the perception of the other. Secondly, congruent events hold greater significance relative to incongruent events and can permit the direction of audience attention. Finally, the identification of temporal congruence can encourage the perception of semantic congruence and vice versa.

3.4 DISCUSSION

Theory regarding cross-modal interaction between sensory stimuli indicates a mechanism for the formation of audiovisual correspondence. Indeed, the phenomena of spatial ventriloquism, temporal ventriloquism and inter-sensory interference highlight the importance of audiovisual synchrony within the formation of cross-modal correspondences. Further to this, studies regarding the phenomena indicate that cross-modal events presented within a certain window of asynchrony are generally integrated despite significant spatial disparity. The application of this concept to the field of audiovisual art implies that, within a simple audiovisual scene, temporal synchrony takes precedence over spatial synchrony. In addition, cross-modal correspondence may be more clearly perceivable between medium events that are presented in synchrony than those which are asynchronous.

Perhaps the most significant implication of theory regarding cross-modal interaction is the role of temporal structure in the process of audiovisual perception. Research indicates that correlations between medium structures will typically result in the perception of a generalised cross-modal congruence. Assuming that correlations between the fundamental temporal patterns remain discernible, this may be derived despite elements of asynchrony or superfluous information (Lipscomb and Kendall, 1994). Further to this, temporal congruence can also be established through equivalent cross-modal transformation (Iwamiya and Ozaki, 2004).

This research posits that temporal congruence offers a mechanism for the manipulation of cross-modal correspondence within the context of audiovisual art. Indeed, literature indicates that the perceived relationship between medium accent structures is often the first stage in a cognitive pairing of audiovisual objects (Cohen, 2001) and encourages further evaluation of the audiovisual scene (Iwamiya and Ozaki, 2004). Similarly, temporal congruence facilitates the direction of audience attention within a complex audiovisual scene. For example, when presented with an auditory event of particular structure, audience attention will be directed to a temporally congruent visual event if one exists (Marshall and Cohen, 1988; Cohen 1998). These observations are fundamental to the composition methodology presented in Chapters 4 and 6 and implemented in the practical component of this research.

Research indicates that the evaluation of semantic congruence involves the analysis of medium associations based on culturally-learned knowledge (Cohen, 1993; Marshall and Cohen, 1988; Lipscomb and Kendall, 1994). Within audiovisual works dealing with referential material this process is simple: media pairings that refer to the same source object will appear more congruent than those which imply abstract connections or subvert established referential knowledge. For media pairings dealing with abstract material in either or both mediums, however, semantic associations are more complex, requiring additional information or explanation. Within this context, media pairings derived from metaphor or linguistic convention are likely to exhibit greater semantic congruence than those with no cultural basis (Coulter, 2010). It is interesting to note that the perception of semantic congruence may be encouraged through conditioned response to temporally congruent events. For example, it is logical to assume that two synchronised but semantically incongruent events may, in time, be deemed semantically congruent if repeated in a consistent manner (Cohen, 1993). This research considers semantic congruence to be an important tool in the creation of abstract audiovisual art that may be defined by thematic elements, such as the mood or tone of the piece, or perceptually equivalent properties within each medium.

3.5 CONCLUSION

In conclusion, theory regarding audiovisual integration and temporal congruence highlights the importance of temporal synchrony as a mechanism for the creation of correspondences between medium structures. On the basis that temporal congruence encourages increased evaluation of the audiovisual scene, this research considers structural correspondences to be a fundamental element in the composition of abstract audiovisual media. Further discussion of the application and effects of temporal congruence within this context is presented in the following chapter. Theory regarding semantic congruence indicates the potential for cross-modal relationships derived from perceptually equivalent properties of each medium. It is the perspective of this research that semantic associations between medium events support and enhance correspondences derived from temporal structure. As such, further discussion of this concept with specific relevance to audiovisual density is presented in Chapters 5 and 6 and is explored within the practical elements of this research.

4 Audiovisual Time

A Formalist Approach to Audiovisual Composition

4.1 INTRODUCTION

Within this chapter, the concept of time as a unifying factor between aural and visual mediums forms the basis for the discussion of both a multi-scale approach to audiovisual composition and the potential for temporal interaction in the resultant media. The chapter begins with an overview of the electroacoustic idiom about which much of this thesis is structured. This is followed by a review of the structural levels for musical organisation across multiple time scales and the discussion of temporal morphology within this context. The concept of visual morphology as a mechanism for the composition and analysis of visual works is then discussed as a mechanism for the identification of temporal congruence between medium structures. Prevalent formalist theory is then reviewed with regards to temporal congruence, permitting the discussion of suspense, expectation and signposting within abstract audiovisual composition.

4.2 THE ELECTROACOUSTIC IDIOM

This research argues that the creation of abstract visual media can be informed by composition methodologies prevalent within the field of electroacoustic composition. Central to this thesis is Schaeffer's concept of the 'sound object.' Described by Chion (1983:32) as 'a sound unit perceived in its material, its particular texture, its own qualities and perceptual dimensions,' the sound object is the manifestation of Schaeffer's theory of *epoché*: the 'methodological process of reduction' of sound perception to only its intrinsic properties (Kane, 2007:3). This focus on the formal properties of sound allows the electroacoustic composer to develop a language of sonic objects that may be classified and organised based on the temporal variance of their spectral properties.

An equivalent phenomenological approach can be adopted in the process of visual composition. Garro (2005:2) observes that 'attention to spectral properties [and] discourse based on the articulation of morphological attributes … are concerns that can be transported into the visual media.' On this basis, a visual equivalent to Schaeffer's sonic object can be defined, encompassing visual properties such as form, texture, colour and spatialisation. As with their sonic equivalent, a taxonomy of 'visual objects' can also be developed, permitting their classification and organisation based on formal properties. Within this compositional paradigm, Garro (2005:3) notes the importance of temporal morphology within both sonic and visual constructs, observing that 'as the attributes of both sound and visual change in time, idiomatic characteristics may emerge specifically from the practice of articulating the time-varying phenomenological attributes.' This approach to composition is discussed by Piché (Steenhuisen, 2009:262), who notes that during the creation of *Sieves* (2004), composing with image in the same manner as sound ensures that ‘the complexity of the image is associated with the complexity of the sound.’ Further to this, Piché (*ibid.*) notes the importance of temporal synchronisation of medium properties as ‘an obvious way of having the parts interact.’ This observation reinforces the position of this thesis: equivalent composition methodologies can be adopted for the manipulation of formal properties in both aural and visual mediums, with the time-variant behaviour of such properties forming a unifying factor in the perception of each.

4.3 TEMPORAL STRUCTURES OF AUDIO AND VIDEO

In his text ‘... How time passes ...’ (1957) Stockhausen presents ‘a unified view of the relationship between the various time scales of musical structure,’ allowing pitch and rhythm to ‘be considered as one and the same phenomenon, differing only in their respective time scales’ (Roads, 2004:73). This concept of temporal relativity and time-scale within musical composition leads to the discussion of a ‘multi-scale’ (Vaggione, 1996) approach to musical composition in which temporal structure may be divided into discrete organisational levels that exist within specific time-scales.

The smallest relevant time scale is the micro-level. Ranging from several hundred microseconds to around one hundred milliseconds in duration, this level comprises ‘transient audio phenomena’ (Roads, 2004:20). Heard in isolation, micro-level sound events are perceived as clicks, but when occurring in rapid succession at a rate greater than around 20Hz, they are perceived as a continuous tone (Wishart, 1996; Roads, 2004). Sonic events with duration between 100ms and a few seconds are termed heterogeneous sound objects and exist within the mini-structural or sound-object organisational level (Xenakis and Brown, 1989; Blackwell and Young, 2004; Roads 2004). Wishart (1996) argues that such structures may discard the static homogeneous characteristics of traditional musical note, instead incorporating time-variant properties to facilitate a ‘dynamic morphology’ of object structure. Sonic events of greater duration, typically measured in seconds, exist within the meso time-scale. Such structures can be textural in nature (Roads, 2004), gestural (Wishart, 1996), or can be the combinatorial product of numerous sound object-level structures (Blackwell and Young, 2004). The subsequent organisation of meso-level structures into musical form, typically minutes in duration, makes up the overall macrostructure of the composition. Roads (2004) notes that without explicit description of musical form prior to performance, the macro time scale may only be perceived retrospectively.

Conceptually, any sonic object within any organisation level that incorporates time-variant properties can be said to exhibit a temporal morphology. Perceptually, however, the strong interaction between time-scales provides for the circulation of morphologies from one level to another. As a result, morphologies realised within a particular organisational level may not necessarily be perceived within the same time scale (Roads, 2004). Indeed, from the perspective of an audience, it is likely that the cumulative product of morphologies within each level would be perceived rather than the individual morphologies themselves. Acknowledging this inherent interaction between organisation levels, Vaggione (2001) observes that a multi-scale approach to music composition requires structural morphologies to be validated not only by their intended effect, but also by their perceivable effects across other time scales.

Returning to the electroacoustic idiom discussed previously, it is reasonable to assume that if visual material can be composed in a manner analogous to the composition of music, then a multi-scale composition methodology can theoretically be applied to the process of visual composition. Any discussion of temporal morphology within such an approach must, however, acknowledge potential disparities between the time-scales of the auditory and visual mediums (Courribet, 2007). Indeed, the translation of compositional strategy from the auditory to visual domain must remain highly conceptual, as a literal transposition of methodology from one medium to the other will often by undesirable and highly impractical (Wishart, 1986).

This research posits that the visual scene can be analysed in terms of time-variant properties, discarding any extrinsic or referential information from which the imagery is derived. Within this context, formal properties of the visual scene can be discussed in terms of temporal morphology within each structural level. Although not explicitly stated as the composer's intention, this is demonstrated in Bret Battey's *Sinus Aestum* (2009), in which the high-level morphologies of the many visual elements contribute to multiple meso-level morphologies in visual density and form. The organisation of these gestures within the macro-structure of the work results in undulating variations in visual form that complement the shifts in spectral density present within the soundtrack.

Acknowledging Wishart's (1996:93) observation that sound objects should be discussed as 'totalities … with various properties, rather than as collections of parameters,' the term 'property' within the context of visual morphology is deliberately vague. On one hand, a property can be an explicit quantitative value derived through mathematical or statistical processes, whilst on the other it can be qualitative, derived from the subjective analysis of perceptual qualities. As such, properties need not be explicitly defined prior to analysis. Indeed, in many cases it is the perception of change, or morphology, that permits their identification. Translating this concept to the practice of visual composition, this research argues the existence of a multi-scale approach to composition that defines form as a series of modulations in visual parameters within each organisational level, irrespective of the referential basis from which the visual material is derived.

4.4 SYNCHRESIS

While a multi-scale approach to audiovisual composition offers a methodology for the creation of audiovisual structures, perhaps of greater importance to the composer is the potential for interactions between the temporal structures of each medium. Indeed, as noted by Piché

(Steenhuisen, 2009:262), 'the aesthetic experience comes from recognizing that there are conjunctions.' One mechanism for this is the phenomenon of *synchresis*. Discussed by Chion (1994:63), synchresis is the point in an 'audio visual sequence during which a sound event and a visual event meet in synchrony,' that acts as a 'spontaneous and irresistible weld' between medium events. Chion (*ibid.*) argues that this phenomenon is the product of not only conditioned response, but also the gestalt laws of grouping. This allows synchresis to function 'with images and sounds that strictly speaking have nothing to do with each other.'

To illustrate the potency of this effect, Bailey *et al* (2007) offer the example of footsteps (an extension of a similar example provided by Chion). Within this, it is observed that our culturally derived expectation that the visual effect of walking will be accompanied by a temporally equivalent audio component, permits great flexibility with regards to the actual audio that is used. Indeed, Bailey *et al* (*ibid.*) go on to offer that if the audio component consisted solely of a sine tone with an amplitude envelope equivalent to the motion of the video, synchresis would still occur. This observation corresponds with Chion's (1994) argument that the response is innate, but culturally reinforced.

It should be noted, however, that there is potential for the overuse of synchresis within audiovisual works. Indeed, compositions that rely heavily on the phenomenon may rapidly cease to be interesting as the connection between medium morphologies becomes less challenging (Dannenberg, 2005). In discussing the practical application of synchresis, Piché (Steenhuisen, 2009:262; see also Piché, 2003) notes that while 'synchronization points are an obvious way of having the parts interact … those moments will typically only compose 10 to 20% of the work.' Instead the use of synchresis should be analogous to 'a chord progression [that] leads to a resolution on a tonal centre.'

4.5 TEMPORAL CONGRUENCE

Presented in Chapter 3, temporal congruence is a qualitative measure of the cross-modal similarity of temporal structures within audiovisual media. It is perceived despite elements of asynchrony and is a means of manipulating audience attention towards structurally similar events (Marshall and Cohen, 1988; Lipscomb and Kendall, 1994; Cohen, 1998; Cohen 2001; Lipscomb and Tolchinsky, 2004). If the temporal structure of both auditory and visual mediums can be defined as a series of morphologies within each organisational level, then it is reasonable to conclude that perceptual correspondences between the morphologies of each medium enable the formation of temporal congruence within the resultant audiovisual output. Within this context it is feasible that if an auditory property and a visual property share a similar time-variant behaviour, or morphology, then the medium events from which they are identified will appear more congruent than events that exhibit heterogeneous morphologies. The formation of temporal congruence is not dependent on an explicit synchronisation of medium events. Instead it is the perception of relative equivalence between each temporal structure that permits the identification of congruence (see Chapter 3). For example, equivalent morphologies in both the size of a circle and the pitch of a sine wave may be deemed congruent despite differences in parameter values as the relative change remains constant. As such, this research considers that it is the relative state of temporal congruence between mediums that permits the interpretation of higher-level emotion or meaning within abstract audiovisual media.

In his seminal text *Emotion and Meaning in Music*, Meyer (1956:3) establishes a formalist approach to musical analysis as one in which the 'meaning of music lies in the perception and understanding of the musical relationships set forth in the work of art.' This position does not explicitly disregard the value of referential connotations within music, instead highlighting the value of 'intramusical' properties within the perception of meaning (*ibid.*:20). Central to Meyer's thesis is the view that 'emotion or affect is aroused when a tendency to respond is arrested or inhibited' (*ibid.*:14). This, he argues, rises out of an innate tendency to seek the resolution of 'doubt and confusion' (*ibid.*:16). This concept forms the basis for the theories of *suspense* and *expectation*. Meyer argues that 'suspense is essentially the product of ignorance as to the future course of events' and gives rise to both 'apprehension and anxiety.' Musical expectation is the product of 'sound successions common to a culture, a style, or a particular work,' deviations from which result in inhibition and therefore emotional affect (*ibid.*:27-32). As noted by Meyer (*ibid.*:32), 'the customary or expected progression of sounds can be considered as a norm, which from a stylistic point of view it is; and alteration in the expected progression can be considered a deviation. Hence deviations can be regarded as emotional or affective stimuli.' While there is clearly significant scope for the application of such theory to the general field of audiovisual composition, of particular interest to this research is the relevance of suspense and expectation in regards to the relationship between the temporal structures of each medium.

Cazden (1945:5) observes that 'the critical determinant of consonance or dissonance is expectation of movement,' with 'a dissonant interval [causing] a restless expectation or resolution, or movement to a consonant interval.' On this basis, this research considers that the perceptual congruence of temporal morphologies may be deemed similar to tonal consonance within music, in terms of emotional effect. Derived from Evans' (2005) discussion of consonance within visual composition and supported by Meyer's (1956) theory regarding expectation, it is reasonable to suggest that incongruent medium structures can create suspense and an expected resolution to congruence, analogous to the resolution of tonal dissonance within music. For example in *Permutations* (1968) by John Whitney, the medium morphologies develop from a state of incongruence (5:20-5:35) to congruence (5:35-5:58), with the rapid transition between states adding greater impact to the change in both aural and visual form. On this basis, this thesis asserts that audiovisual works which exhibit temporal dynamism in the perceptual congruence of formal structures may present to the audience a more challenging and therefore emotionally affective experience.

Another potential application of formalist theory to the field of audiovisual composition is the concept of audiovisual signposting: the expectation of an audiovisual event based on the perception of another. Meyer (1956:35) refers to this concept as *embodied meaning*, noting that 'what a musical stimulus or a series of stimuli indicate and point to are not extramusical concepts and objects but other musical events which are about to happen.' As such, a musical event 'has meaning because it points to and makes us expect another musical event.' Translating this theory to audiovisual structure, it is reasonable to presume that a particular relationship between medium morphologies may encourage the expectation of another based on the events that have preceded it. It is this expectation that gives the audiovisual gesture embodied meaning. Perhaps more interesting however, is the potential for emotional arousal through the subversion of expectation. To this end, the contradiction or modulation of pre-existing temporal relationships may prove more emotionally affective than the simply conforming to expectation.

From the perspective of audiovisual composition, all but the simplest works will exhibit multiple morphologies within each medium. Returning to the previous example, the colour of the circle may also exhibit a temporal morphology as well as the size. These two visual morphologies may develop in a relative manner or independently. The combinatorial behaviour of multiple morphologies is analogous to the concept of musical polyphony, with relative morphologies appearing monophonic and independent morphologies polyphonic. From the perspective of an audience, however, the analysis of structural congruence need not be performed on a 'per-voice' level. Instead, it is the combination of multiple temporal structures within each medium and the cross-analysis of the combined product that allows the identification of perceptual congruence between morphologies. It is likely that such structural complexity will have a strong effect on the perceived temporal congruence. For example, two equivalent morphologies will typically be deemed more congruent if displayed in isolation than in a complex, polyphonic audiovisual scene, a concept alluded to by Dannenberg (2005). On the basis that the relative polyphony of each medium will likely be a factor in the evaluation of temporal congruence, it is also probable that any increase in temporal complexity in one medium will require an equivalent increase in complexity in the other to maintain congruence.

While the re-use of structural relationships may prove an effective technique to increase audience expectation towards a certain resolution, it is probable that through repetition, standardisation will occur and the emotional effect of such structures will be reduced. Meyer (1956:68-69) presents two processes by which standardisation can be negated in musical composition. Firstly, 'the degree of deviation can be increased.' Within the context of audiovisual structure, this would encourage either the degree of relative congruence or incongruence to be increased, or a modulation in temporal complexity to add further detail to the standardised event. Typically this modulation would occur within the micro or sound-object levels. Secondly, Meyer (*ibid.*) suggests that 'new deviant devices can be introduced in the style as alternatives, weakening the probability relationships between the former deviant and its consequents.’ Within the context of audiovisual composition, this could involve the meso or macro-level introduction of alternative temporal structures, with varying congruence and complexity. This would then encourage a re-evaluation of the expected temporal progression and therefore increased emotional affect as the audience seeks resolution for the structural change.

4.6 A MODEL FOR AUDIOVISUAL COMPOSITION

Derived from the presented theory, a conceptual model for audiovisual composition can be established. When working with abstract or non-referential materials, the structure of audiovisual media can be defined as a series of time-variant properties to which temporal morphologies may be assigned. Such morphologies can be applied within each organisation level, ranging from the micro-level time scale to the overall macro-structure of the composition. Similarly, the morphologies of identified properties may be homogeneous, resulting in a monophonic temporal structure, or heterogeneous, resulting in a polyphonic temporal structure. The cross-analysis of medium structures at any organisation level permits the identification of temporal congruence, the modulation of which allows the creation of a dynamic relationship between temporal structures. It is this measure of congruence that enables the perception of a higher level emotion or meaning within the work.

While this model offers a conceptual framework for audiovisual composition, the perception of temporal congruence remains subjective and the audience perception of temporal structures may differ from those intended by the composer. The primary reason for this disparity is the perception of morphologies within multi-scale composition. As discussed previously, there exists an inherent interaction between morphologies within each organisational level (Vaggione, 2001; Roads, 2004). On this basis, it is unlikely that an audience would be aware of low-level morphologies and resultant temporal structure without prior explanation of the compositional process. Instead, the combinatorial result of numerous morphologies within the composition meso or macro-structure would typically be perceived. This disparity between composer intention and audience perception highlights the limitations of the model as a mechanism for the creation of objective correspondences between aural and visual mediums.

4.7 CONCLUSION

In conclusion, theory regarding the application of electroacoustic composition techniques to the field of visual art has been identified as a basis for the discussion of a multi-scale approach to audiovisual composition. From this, the concept of audiovisual parameters as semantic categories for the implementation of temporal morphology has been discussed, with perceptual correspondence of such morphologies permitting the analysis of temporal congruence. Finally, modulation in the relative state of temporal congruence has been identified as a mechanism for the creation of suspense and expectation in a manner similar to the resolution of dissonance in tonal music. These conclusions form the basis for a conceptual model for audiovisual composition. The practical application of this is presented as in Chapter 6 as a framework for the creation of abstract audiovisual art.

5 Parameter Mapping

5.1 INTRODUCTION

Literature from the field of musical instrument design offers a wealth of terminology and techniques for the creation of parameter mapping strategies for controlling audio synthesis. Conversely, while the potential role of parameter mapping within audiovisual art is acknowledged, explicit discussion of its integration within composition methodology is rare. This chapter explores parameter mapping literature from the fields of instrument design, algorithmic composition and audiovisual art to allow mapping strategy within the context of audiovisual composition to be discussed. Following the identification of prevalent parameter mapping terminology, the concept of mapping hierarchy is presented. The issue of mapping transparency is then evaluated in both general terms and within the context of audiovisual composition. The chapter concludes with a discussion of methods for the manipulation of transparency within audiovisual mapping strategies.

5.2 MAPPING TYPES

Hunt *et al.* (2000; see also: Hunt and Wanderley, 2002) propose that two forms of mapping exist: those that use generative mechanisms to perform mapping and those that use explicitly defined mapping strategies. Within the latter category, Arfib *et al.* (2002:130) offer a distinction between ‘explicit mappings with mathematical expressions clearly defined for each mapping connection, and explicit mappings with a general rule.’ On this basis *explicit* mappings are defined as those in which ‘one can exactly describe the links between the input and the output parameters,’ while *implicit* mappings are defined as those in which ‘the mapping box is considered as a black box for which we define behaviour rules but not precise value rules’ (*ibid.*). The use of explicit and implicit mapping strategies is also discussed within the context of algorithmic composition by Doornbsuch (2002). Throughout this research the term implicit is adopted to describe the mapping of parameters derived from the perceptual analysis of one data set to another. Within this context, the perception of change in one parameter can be used to inform relative change in another without the need for a direct translation of actual values. Conversely, the term explicit is adopted as a descriptor for the mapping of literal values between data sets, incorporating mathematical functions as required.

A distinction can also be drawn between *static* and *dynamic* mappings. In the former, the relationship between input and output parameters remains constant, whilst in the latter the relationship between parameters is variant based on pre-defined gestural data, or through adaptation to input data (Arfib *et al.*, 2002; Bevilaqua *et al.*, 2005; Momeni and Henry, 2006).

Within parameter mapping literature the terms *linear* and *non-linear* are often adopted as alternative descriptors for static and dynamic mappings respectively (Hunt *et al.*, 2000; Doornbusch, 2002). As such, within an audiovisual mapping strategy the terms static/dynamic and linear/non-linear can be used interchangeably as descriptors for the time variant behaviour of a mapping strategy.

Mappings can be classified as one of three basic types: *one-to-one*, *one-to-many* and *many-to-one*. Through combination it is also possible to create a *many-to-many* parameter mapping strategy (Hunt *et al.*, 2000; Arfib *et al.*, 2002; Fels *et al.*, 2002). Further to this, one-to-one mappings may be described as *simple* mappings, whilst one-to-many, many-to-one and many-to-many mappings may be defined as *complex* (Hunt and Wandereley, 2002). Similarly, Rovan *et al.* (1997) propose the terms *divergent* and *convergent* as descriptors for one-to-many and many-to-one mapping strategies respectively. Doornbusch (2002:155) defines simple mappings as those which are 'ratiometric,' in which 'a doubling of the data to be mapped results in a "doubling" of the musical parameter.' Conversely, complex mappings are defined as those which may be 'regularly non-linear' or those that 'obscure the original concept, compositional gesture or data.' For the purposes of this research, the explicit definition of mapping complexity prevalent within interface design literature is adopted. It should be noted, however, that this definition is independent from more generalised discussions of mapping complexity which refer to the overall relationship between input and output data rather than individual mappings employed (Doornbusch, 2002; Dannenberg, 2005).

5.3 MAPPING HIERARCHY

Within parameter mapping literature, the term *layer* is used to describe an independent mapping system within the context of a parameter mapping strategy. Hunt *et al.* (2000; see also: Hunt and Wanderley, 2002) present a multi-layer parameter mapping model for the translation of performance data to synthesis parameters, with similar approaches being documented in Arfib *et al.* (2002), Mulder *et al.* (1997), Hunt *et al.* (2002) and Rovan *et al.* (1997). Within the context of real-time media generation Momeni and Henry (2006) discuss the relevance of both 'classic' and dynamic mapping layers to facilitate concurrent audiovisual synthesis derived from performance parameters. Mapping layers within an encompassing strategy may exhibit independent complexity and dynamism. Similarly, it is possible to incorporate both implicit and explicit mapping layers within a single mapping strategy, although there are no direct examples of this within parameter mapping literature.

Multi-layered parameter mapping strategies can be described as either *serial* or *parallel* (Figure 5.3.1). In the former, the relationship between input and output parameters is a result of the cumulative effect of multiple mapping layers and permits the translation of data between two discrete sets, most clearly demonstrated in Arfib *et al.* (2002). Conversely, a parallel mapping strategy allows the dissemination of data from a single input set to multiple output parameter sets, with output data being subject only to the mapping layer between itself and the input data set, as demonstrated by Momeni and Henry (2006). Through the combination of serial and parallel mapping hierarchies, mapping strategies of varying complexity can be created, facilitating the use of multiple input and output data sets.

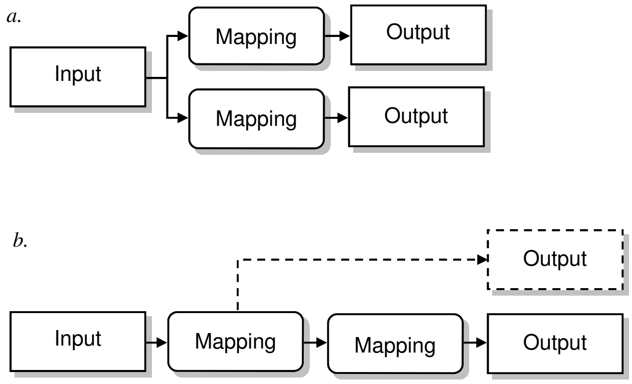

Figure 5.3.1 – A flowchart illustrating parallel (a) and serial (b) mapping hierarchies.

This research finds that mapping hierarchies within audiovisual composition generally fall into one of two categories: *media-driven* or *data-driven*. Within this categorisation, a distinction is drawn between data derived from either aural or visual mediums (media-driven) or data derived from an external source (data-driven). This distinction is made on the basis that in the former both input and output data are presented within the final composition, while in the latter only the resultant output can be perceived directly. In a media-driven mapping hierarchy, parameters derived from the perceptual or computational analysis of one medium can be translated to the other by means of implicit or explicit mapping functions respectively. A notable example of this methodology within the field of audiovisual composition is Adrianno Abaddo's *Dynamics* (1988) which relies on the perceptual mapping of audio features to the visual domain (Abbado, 1988). Similarly automated visualisation systems prevalent within VJ practice rely on the analysis of audio data for the creation of visual material (Alexander and Collins, 2008). In a data-driven mapping hierarchy a single data set is mapped to both aural and visual mediums to facilitate concurrent audiovisual synthesis. Typically the employed external data set consists of gestural or human interface data as discussed by Edmonds *et al.* (2004) and Momeni and Henry (2006), although mappings may also be derived from the statistical analysis of real-world data as in *Concrete Net* (1996) by Jøran Rudi (1998, 2005). It should be noted, however, that mapping strategies within audiovisual composition need not strictly adhere to either hierarchy and that multiple data-sets can be employed, dependent on the conceptual and aesthetic direction of the work.

5.4 MAPPING PERCEPTION

In discussing expressivity within musical instrument design, Fels *et al.* (2002:109) define the term *transparency* as 'an indication of the psychophysiological distance, in the minds of the player and the audience, between the input and output of a device mapping.' Extrapolating from this, this research considers a generalised definition of parameter mapping transparency as a qualitative measure for the perceivable relationship between input and output parameters. Within this definition, a mapping in which the relationship between input and output parameters is clearly perceivable is described as transparent, while a mapping in which the relationship between input and output parameters is imperceptible is described as opaque. This terminology can be applied to the practice of media-driven audiovisual art, specifying the strength of any cross-modal correspondences that result from the adopted mapping strategy. Within a data-driven mapping hierarchy, transparency specifies the perceptual strength of correspondences between aural and visual events and also between input data and the resultant medium structures.

Discussing parameter mapping within audiovisual composition, Dannenberg (2005:28) observes that ‘as soon as the obvious connections from sound to image are made, the image ceases to be interesting or challenging,’ noting that by using mappings that are not clearly apparent ‘the audience may perceive that there is some emotional, expressive or abstract connection, but the animation and music can otherwise be quite independent and perhaps more interesting.’ Within the field of instrument design, complex mappings have proven more engaging for the performer (Hunt and Kirk, 2000), an observation that Doornbusch (2002) postulates can translate to the field of algorithmic composition. By implication, transparent parameter relationships resulting from a simple mapping strategy may result in an audiovisual work that lacks depth and interest (Alexander and Collins, 2008). The sole use of opaque mappings would, however, be inappropriate for compositions that seek to define an interpretable correspondence between mediums. Instead, practical experimentation throughout this research finds that audiovisual composition is best served by parameter mapping strategies that exhibit temporal variance in mapping transparency, encouraging modulations in cross-modal correspondence analogous to the use of consonance and dissonance within tonal music or the creation of musical counterpoint and harmony (see Chapter 4).

While the concept of mapping transparency offers terminology to facilitate the discussion of parameter mapping strategy, of greater importance to the process of audiovisual composition is the identification of techniques for its manipulation. Dannenberg (2005:29) observes that ‘if a particular sound or gesture affects some visual parameter, the audience is much more likely to “get it” if the sound and the visual effect are presented in relative isolation’ noting that if a cross-modal relationship ‘takes place amid a chaotic blend of other sounds and images, it is unlikely anyone will detect the connection.’ Extrapolating from this, this research finds that complex mapping strategies, such as those that employ dynamic or complex mappings, will typically result in a more opaque relationship between input and output parameters relative to less complex mapping strategies such as those that employ static or simple mappings.

Mapping transparency can also be controlled through the modulation of audiovisual synchrony. Typically, a parameter mapping strategy will rely on the temporal correlation of input and output parameters to ensure mapping transparency. Indeed, Mulder *et al.* (1997) note that the presence of temporal latency between human interaction and audio output within instrument design, results in ineffective mappings that can appear counter-intuitive or opaque. Acknowledging the effects of temporal ventriloquism upon audiovisual integration (Spence and Squire, 2003; Sugita and Suzuki, 2003), it is probable that the introduction of latent or asynchronous processes within an audiovisual mapping strategy may offer a mechanism for the modulation of mapping transparency, with asynchronous events having greater opacity than those which occur in synchrony.

Fels *et al.* (2002:109) identify metaphor as 'one technique to facilitate moving from an opaque mapping to a transparent mapping' by enabling 'device designers, players and audience to refer to elements that are "common knowledge."' Indeed, Alves (2005:47) argues that the 'literal mapping of pitch space to height [is] only intuitive because our culture has adopted that particular arbitrary metaphor of "low" and "high" to describe pitch,' whilst Coulter (2010:29) notes that 'if commonly used metaphors are produced when contemplating the referential content of [audiovisual] materials, then a simple association of audio and video media should be sufficient to transmit the creative idea.' Such observations are supported by theory from the field of audiovisual integration (see Chapter 3), in which it observed that semantically congruent media pairings enhance audiovisual integration relative to those which are semantically incongruent (Molholm *et al.*, 2004). Consequently, this research posits that audiovisual mappings that are derived from metaphor will prove more transparent than those which are non-referential or abstract. Similarly, mappings that subvert audience expectation by disregarding established metaphoric connections will likely be perceived to have greater opacity relative to those which embrace such relationships. The role of metaphor within audiovisual parameter mapping also raises the question of the cultural subjectivity, with mapping transparency differing amongst individuals dependent on cultural background. Clearly this consideration is of greater importance to compositions that employ referential media derived from a specific culture (Jones and Nevile, 2005; Coulter, 2010).

5.5 DISCUSSION

Individual parameter relationships within an audiovisual mapping strategy can be discussed in terms of transparency as a measure of the perceptual bonding between input and output data. The perceptual transparency of each mapping will be highly personal and may be affected by factors such as cultural background and pre-existing knowledge. For example, if an audience is aware of the fundamental mapping structure employed throughout the creation of an audiovisual work, it is probable that the parameter relationships will be more apparent throughout its performance. Similarly, metaphoric mappings that are culturally derived may appear transparent to an audience from the culture concerned, despite being opaque to an audience from a different cultural background. The transparency of a mapping can also be controlled through modulations in scene-complexity and audiovisual synchrony. For example, an asynchronous mapping between visual height and audio pitch in a complex audiovisual scene may appear more opaque than the same mapping occurring in isolation and synchrony, regardless of the metaphoric nature of the relationship. Within the context of this research, transparency provides terminology for the analysis of parameter mappings within both media-driven and data-driven mapping strategies. Additionally, this research finds that the modulation of perceived mapping transparency can offer a mechanism for complex cross-modal relationships that negate the inherent predictability of strongly mapped material.

Whilst the identified terminology permits the analysis of parameter mapping strategy, of greater importance to the audiovisual composer is the role of parameter mapping within the composition methodology and the intended transparency of parameter relationships within the resultant media. Within the context of algorithmic composition, the responses of the composers interviewed by Doornbusch (2002) consistently indicate that the mapping process need not be objectivised within the resultant composition. Indeed, none of the composers claim the perception of mappings between data and music to be the aim of their works, a distinction that this research considers should be observed within the context of audiovisual composition. On this basis, the use or discussion of parameter mapping within audiovisual composition should acknowledge the intended role of such mappings throughout the composition process.

Within the context of media-driven audiovisual composition, this research finds that parameter mapping offers a mechanism for the translation of data from one medium to the other, with musical structure being derived from the visual morphology or vice-versa. An explicit approach to mapping within this context involves the extraction of values from the dominant medium, a process of translation using mathematical functions and the application of the results to the dependent medium. An example of this approach would be the literal mapping of visual luminosity to aural amplitude. Conversely, an implicit approach to mapping requires the identification of perceived morphologies within the dominant medium and the application of this structural data to the dependent medium through a process of artistic interpretation rather than through the use of explicit data values. The use of perceived visual density as a structural basis for the modulation of audio spectral density would be an example of an implicit mapping within media-driven audiovisual composition.

Within a data-driven approach to audiovisual composition, parameter mapping facilitates the creation of audiovisual structure through the translation of values from an external data-source to both aural and visual domains. In this context, an explicit mapping process involves the literal translation of values from the selected data-set to each medium, using mathematical functions as required. An example of this methodology would be the use of values derived from simulation data, such as cellular automata or flocking simulation, to simultaneously control parameters within the both visual and aural domains, such as colour and pitch. Conversely, an implicit approach to data-driven audiovisual composition involves the translation of perceived properties within the source data into aural and visual structure. For example, the subjective properties of real-world objects as interpreted by the composer may be applied to parameters within each medium, usually involving metaphoric correspondences. Clearly, an implicit approach to data-driven audiovisual composition requires significant abstraction between source-data and audiovisual output and it is unlikely that such an approach would be discussed as an explicit stage of composition practice. However, this research considers the use of implicit mappings within data-driven audiovisual composition to be a valid methodology to inform the generation of audiovisual structure. Further discussion of data-driven approaches to audiovisual composition is presented in Chapter 6.

5.6 CONCLUSION

In summary, a theoretical basis for the creation of parameter mapping strategies has been offered with specific relevance to the field of audiovisual art. Relevant terminology prevalent within the fields of musical instrument design and algorithmic composition has been identified as a mechanism for the creation and analysis of parameter mapping strategies within the context of audiovisual composition. Similarly, the terms media-driven and data-driven have been identified as descriptors for audiovisual mapping hierarchy. Transparency has been identified as terminology to define the relationship between input and output data. Further to this, audiovisual complexity, synchronisation and semantic unity have been identified as primary mechanisms for the manipulation of transparency within an audiovisual mapping strategy. Finally, the role of parameter mapping within audiovisual composition has been discussed in order to inform the practical application of such theory within the creation of audiovisual art. Further discussion of parameter mapping theory within the context of audiovisual composition methodology is presented in the following chapter.

6 Audiovisual Particles

Parameter Mapping as a Framework for Audiovisual Composition

6.1 INTRODUCTION

The primary aim of this chapter is to describe the framework for audiovisual composition developed throughout this research. Informed by practical experimentation with multi-scale particle organisation techniques, a conceptual parallel between audio grains and visual particles is identified, the concurrent synthesis of which is presented as a mechanism for the creation of temporal congruence. Referencing the field of electroacoustic composition, density is offered as terminology to inform the composition process and to permit the analysis of particle structures. Following this, the concept of data-driven composition is presented as a methodology for the transposition of morphological processes and formation of temporal congruence. Finally, a framework for audiovisual composition is offered as a practical application of theory discussed in Chapters 4 and 5. The presented framework defines a methodology for the creation and organisation of audiovisual gesture and the manipulation of temporal congruence within audiovisual media.

6.2 AUDIOVISUAL PARTICLES

Serafin (2007:207) defines a grain of sound as ‘a short sonic snippet of about ten to a hundred milliseconds, an elementary particle as opposed to a complex soundscape.’ The organisation of such ‘transient audio phenomena’ (Roads, 2004:20) into discrete temporal structures permits the formation of sonic ‘clouds’ upon which ‘the composer may impose specific morphologies’ (Roads, 2004:15). Within the context of the organisational levels discussed in Chapter 4, an audio grain is typically subject to morphology within the micro time-scale, while the temporal structure of a grain cloud is subject to variation within the meso time-scale. Granular synthesis is one method for the creation and manipulation of such structures, enabling the algorithmic extraction of micro-level audio grains and their subsequent into larger acoustic events (Roads, 1988, 2004; Truax, 1988, 1990; Tolonen *et al.*, 1998).

Within the field of computer graphics, a particle system can be used to simulate ‘fuzzy’ phenomena such as smoke or fire, allowing for the ‘creation of structure and motion from a relatively brief abstract description’ (Sims, 1990:405). Typically, this structure is represented by a group of particle entities each with an independent ‘state.’ Each state is defined by a set of variables pertaining to either simulation or rendering processes (Krajcevski and Reppy, 2011). The first model for such a system was presented by Reeves (1983) and describes operations for the creation, simulation and rendering of entities within the system. Real-time implementations derived from this model include Sims (1990), Lander (1998), Burg (2000) and McAllister

(2000). More recent GPU-based particle systems include Kipfer *et al.* (2004), Kolb *et al.* (2004) and Latta (2004). Further to this Krajcevski and Reppy (2011) present a particle system library that can be executed on either the CPU or the GPU.

Within the context of 3D computer animation, many software systems currently incorporate particle simulation and rendering capabilities. Notable examples include *nParticles* within *Autodesk Maya* (Pallamar and Keller, 2011), the *Particle Flow* extension for *Autodesk 3ds Max* (Murdock, 2011) and the particle system within the open source *Blender* (Hess, 2007). Similarly, particle system functionality is incorporated within a wide range of real-time rendering engines such as *OGRE* (Junker, 2006), *Unity3D* (Blackman, 2011) and numerous commercial video game engines (Trenholme and Smith, 2008). This prevalence of particle systems within commercial and open source software is indicative of the relevance of particle rendering techniques to the field of computer graphics.

While there exists a strong technological basis for the creation of particle systems, of primary importance to this research is the application of particle technology to the practice of audiovisual composition. Although there is limited discourse regarding the use of particle systems within this context, their use can be seen in many of the works of Jean Piché and Bret Battey (see Chapter 2). Indeed, throughout the creation of *Autarkeia Aggregatum* (2005) and the *Luna Series* of works (2007-09), Battey applied a custom 'Brownian dispersion filter' to the pixels of input imagery, decomposing the image into thousands of autonomous entities each subject to various forces. Similarly, in Piché's *Borreales* (2008-09) and *Australes* (2010-11), data was extracted from source video media and applied to the parameters of a particle generation system, the results of which comprise the visual component of the works. The aesthetic capability of particle systems is perhaps best demonstrated in Piché's later work *Oceanes* (2011). Within this, the combinatorial result of the underlying particle system results in complex, abstract visual structures that exhibit evolutions in both form and texture throughout the duration of the piece.

This research posits that a conceptual parallel can be drawn between the processes of granular synthesis and visual particle systems: both employ algorithmic modulations of the individual parameters of low-level entities in order to create higher level morphologies in structural form. Due to differences in the perception of aural and visual material, it is clear that such a parallel must remain highly conceptual. It stands to reason, however, that if the organisational processes within each system are fundamentally alike, then an equivalent composition methodology can be adopted for each. On this basis, the term *particle* is adopted to describe a low-level aural or visual entity, the organisation of which by synthesis algorithm enables the formation of particle *clouds*, a term derived from Roads (2004). Within the context of this research, it is assumed that an audio particle will be rendered as a sound grain, and a visual particle will be rendered as an object within virtual space to which colour and/or image may be assigned. Further to this, the mapping of variables derived from an external data set, or the analysis of a lead medium to synthesis parameters within each medium can facilitate the formation of not only an interpretable correspondence between temporal morphologies, but also a metaphoric correspondence between the perceptual densities of each medium (see Chapter 5).

6.3 DENSITY CORRELATIONS

While the primacy of temporal congruence over metaphorical correspondences is posited in the formalist methodology presented in Chapter 4, the concurrent synthesis of aural and visual particles enables semantic congruence (see Chapter 3) to be defined through perceptual correlations between medium densities. Derived from Roads (2004), the density of audiovisual particle structures can be defined as the number of particles generated per second. Further to this, the *fill factor* of such structures is defined as the product of the cloud density and particle duration in seconds, with mean values being used when the parameters exhibit temporal variation. Roads (2004:332) observes that, within the context of electroacoustic composition, the density of a structure ‘translates to the ratio of sound to silence,’ and that ‘by controlling density and size of sound particles we have a handle on the quality of sonic opacity.’ In practice, it was found that the perceptual density of an audio grain cloud will also be affected by the sonic structures from which the cloud is formed. For example, a dense cloud of grains extracted from spectrally simple source audio, such as a sine wave, may appear perceptually less dense than a relatively sparse cloud of grains extracted from a complex audio waveform. As such, the distinction between theoretical particle density and the perceptual density of the resultant structure as perceived by an audience should be acknowledged.

Within the field of visual particle systems a similar disparity was observed. While a cloud of visual particles will have a density and fill factor defined by the particle creation rate and duration, the perceptual density of the resultant visual structure will also be affected by the particle size and opacity parameters. As such, the perceptual effects of mappings between density-related parameters in each medium will likely be inconsistent, with resultant morphologies being dependent on parameters exclusive to each synthesis algorithm or the source material from which individual particles are derived. For example, a cloud of audio grains extracted from a simple waveform may appear perceptually less dense than an equivalent cloud of large, opaque particles, despite the density and fill factor properties being identical. On this basis, while the concept of cloud density presents a conceptual correspondence between structures in each medium, the interpretation of such effects by an audience will likely be highly subjective. Indeed, this research proposes that a correlation between the temporal morphology of perceptual densities will be more clearly and consistently perceived by an audience than any literal correspondence between the density or fill factor values of each structure. Such an observation does not, however, detract from the use of the terminology as a mechanism for the conceptual and aesthetic direction of a composition or the cross-analysis of particle structures.

6.4 DATA-DRIVEN COMPOSITION

Detailing solutions for the gestural control of synthesis parameters, Momeni and Henry (2006) explore techniques for the concurrent synthesis of aural and visual media derived from a single set of control parameters. Within this context, Momeni and Henry (*ibid.*) argue that a one-to-many mapping system between input data and synthesis parameters permits complete cross-modal synchrony which can then be abstracted as required. The use of an external data set for micro-level organisation is explored by Blackwell and Young (2004) in *Swarm Granulator*, in which the simulated interactions of swarming particles (Reynolds, 1987; Kennedy and Eberhart, 1995) provides a data-set from which granular synthesis parameters are derived. Similarly, Momeni and Wessell (2003) present *Boids Space*, an implementation of flocking simulation within a live performance instrument.

Within the context of linear audiovisual composition, synthesis parameters may be derived from any external data-set from which a temporal morphology can be extracted. This methodology facilitates the manipulation of medium morphologies at varying levels of abstraction, dependent on the composer's intent. Such external data may take the form of linear scripts, stochastic processes or behavioural simulation, amongst others, with parameter mapping facilitating the translation of data between contexts. As such, a generalised model for data-driven audiovisual composition has been defined by this research. Within this, parameters are extracted from an external data set and mapped to synthesis engines, permitting the translation of temporal morphologies from data to media as a mechanism for the formation and dissolution of temporal congruence. Figure 6.4.1 illustrates two strategies for the implementation of such a model.

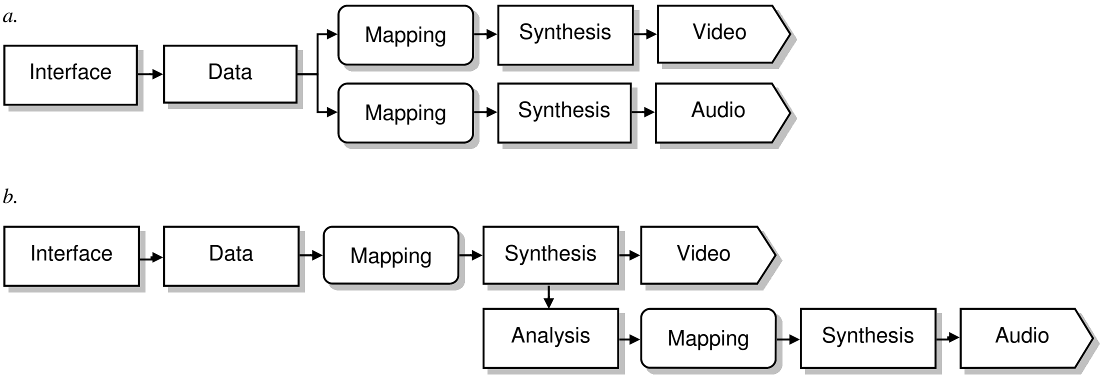

Figure 6.4.1 – Flowcharts illustrating parallel (a) and serial (b) data-driven hierarchies.

Within a parallel data-driven hierarchy parameters are extracted from the source data set and mapped to both audio and video synthesis parameters. Conversely, within a serial hierarchy extracted parameters are mapped to a single synthesis algorithm, with selected output variables being mapped to the second synthesis algorithm following a process of analysis. While correspondences between the data set and second synthesis output will be reduced relative to first, it is likely that morphologies exhibited by the source data will be evident to some extent in both. Within each hierarchy a level of composer interaction is presumed, enabling the modulation of input data and resultant output morphologies. Similarly, the process of parameter selection and the employed mapping system will be defined by the composer and will be specific to each composition. While artist influence upon the resultant audiovisual output will be abstracted as a result, modifications to the external data set permit a level of control over the generated media. As such, the temporal morphologies exhibited within the output will be a result of the interaction between composer and data rather than a product of either component independently.

6.5 A FRAMEWORK FOR AUDIOVISUAL COMPOSITION

The framework developed throughout this research and used in the creation of the compositions presented later in this thesis is described in the following. The adoption of granular synthesis and visual particle systems within this research has been predicated by three arguments. Firstly, data-driven composition techniques permit the translation of temporal morphologies, potentially facilitating the creation of temporal congruence. Secondly, correspondences between perceived medium densities allow the formation of semantic congruence in the resultant audiovisual output. Finally, the use of particle-based composition techniques enables the creation of complex and evolving audiovisual form and has significant potential for abstraction. On this basis, a data-driven framework for the creation of audiovisual media can be defined. While this framework requires parameter mapping to translate data between composer interface, external source and synthesis algorithms, transparent mappings are not intended. Instead, parameter mapping facilitates the transposition of temporal morphology between each framework element. The proposed model is divided into three components: material selection, media generation and media organisation.

The process of material selection specifies both aesthetic and morphological bases from which audiovisual media may be derived. From an aesthetic perspective, the composer must select the source waveforms for granular synthesis and the colour palette, texture images and rendering mechanism for visual particles. The chosen data set defines the morphological basis for the composition at this point, be it behavioural simulation, scripted data or otherwise. The selected materials may be chosen based on semantic correspondence or purely for aesthetic value; while the presented framework facilitates the transposition of temporal morphologies, the artistic basis for the work lies solely with the composer.

Following the selection of source material, the process of media generation involves the synthesis of aural and visual media via either a serial or parallel data-driven hierarchy. Typically this will involve the use of explicit micro or sound-object level mappings to translate data across mediums. The employed parameter mappings should be selected dependent on the conceptual basis for the work and the desired cross-modal correspondences. For example, if a composition seeks to define a perceptual correspondence between medium intensities then relevant synthesis parameters should be mapped to permit this. Conversely, mappings may be purely abstract with no conceptual basis. Such mappings are likely to be imperceptible to an audience, but the transposition of temporal morphologies across mediums will enable the creation of interesting and artistically relevant audiovisual media. The process of media generation is typically interactive to some level. For example, the composer can manipulate the selected data set, or modify the parameter mapping strategy in real-time to impose an alternative morphology upon the resultant media.

In discussing micro-level audio composition, Roads (2004:343) notes the limitations of parameter variation within compositional strategy, observing that composition with a single synthesis algorithm and limited source material may result in not only a ‘monochromatic timbral palette,’ but also ‘a restricted range of gestures, since no morphological developments can occur that are not derived from parameter variation.’ To counteract the potential limitations of such an approach, Roads (*ibid.*) suggests that ‘the compositional strategy can itself be the subject of variations,’ such as an alternative synthesis algorithm or modulations at a different structural level. These observations indicate a potential limitation within strictly mapped audiovisual material. Developing from this, the presented framework incorporates the process of media organisation to permit the modulation of compositional strategy throughout the macrostructure of the work. While the material produced in the process of media generation may be artistically valid in isolation, it is probable that it would rapidly become predictable due to the process of parameter variation from which it is derived. As such, multiple audiovisual gestures should be created employing a range of sonic and visual material and a range of mapping and/or synthesis systems to produce a diverse range of audiovisual material. Typically such gestures would exist within the meso time scale. On this basis, an iterative methodology of media generation is encouraged, in which the results from a particular set of source material, mapping hierarchy and synthesis algorithms are analysed to inform the future source material selections. Such a process enables the creation of a diverse selection of audiovisual media from which the overall composition may be formed.

The macro-level organisation of audiovisual media forms the final stage of the presented framework. Typically the process of media organisation is composer-driven, involving the manual selection and organisation of generated media based on the desired aesthetic and conceptual outcome. Generally, this involves modifications to the temporal structures of each medium to emphasise or reduce temporal congruence to form a dynamic relationship between temporal structures. Dependent on the selected mapping strategy, generated media will typically exhibit varying levels of temporal congruence. For example, transparent mappings between key parameters will result in strong correlations between the temporal morphologies of each medium, whilst more abstract or opaque mappings between lower-level parameters will result in greater incongruence between temporal structures. The composer may impose alternative morphologies upon the generated media to either enhance or abstract the relative congruence of the structures. For example, modifications to macro-level parameters, such as visual brightness or aural amplitude may serve to either reinforce or subvert the pre-existing correspondences between temporal structures. Such modifications can be derived from implicit mappings between the perceptual properties of each medium, or can be purely abstract. Similarly, the generated media need not be presented in the manner in which it was generated. For example, the aural and visual components of a highly congruent gesture may be offset, ranging from out of phase asynchrony to the presentation of each medium in isolation, dissolving any perceptual correspondence between structures. Regardless of the compositional methodology by which it is achieved, the presented framework encourages an approach to audiovisual composition that enables the creation of temporal consonance, dissonance, monophony and polyphony between audiovisual structures.

6.6 CONCLUSION

A literal transposition of theory from the field of electroacoustic micro-composition to the field of visual composition with particle systems is impractical. However, the definitions of audiovisual particle and cloud offer terminology for the discussion of compositional strategy with granular synthesis and visual particle systems, and the theoretical translation of compositional practice between each. On this basis, the presented framework offers a methodology for the concurrent synthesis of aural and visual media and the subsequent organisation of such gestures within audiovisual macrostructure. While the practical application of this may vary dependent on composer intention, the framework offers a foundation for the creation and abstraction of temporal congruence. While the potential for the formation of interpretable cross-modal correspondences is acknowledged, the aim of this theory is to enable the creation of a diverse range of audiovisual material. Further discussion of the practical application of this framework is presented in the following chapter.

7 Composition Studies

7.1 INTRODUCTION

This chapter presents three implementations of the audiovisual particles framework discussed in the previous chapter. While the primary aim of the studies is to inform discussion of the audiovisual particles framework, they are also intended to serve as artworks in their own right. Three works are presented: *Io*, *Rhea* and *Dione*. Each study employs the processes of material selection, media generation and media organisation, with perceptual correlations between medium densities forming the basis for the latter. The process of media generation differs between compositions, with *Io* employing a data-driven, serial approach, *Rhea* employing a data-driven, hybrid approach and *Dione* adopting a process of independent media generation informed by implicit mappings. Following a detailed overview of the adopted composition methodologies, observations relevant to the proposed framework are discussed. The chapter concludes with a discussion of conceptual theory regarding the processes of media generation and organisation. The composition studies are available on the submitted data disc with both source and generated audio material where relevant (see Appendix I).

7.2 IO

The primary aim of *Io* was the exploration of the audiovisual particles framework, focussing on the process of data-driven media generation and the role of parameter mapping within the presented methodology. The source data for the composition was generated by exerting a range of attractive and repulsive forces upon a set of stochastically emitted particles. Force locations were stored within a set of maps that were toggled in real-time, modifying the particle trajectory as a result. The behavioural simulation and visual synthesis systems were developed in *Processing* (Reas and Fry, 2006), permitting the real-time manipulation of simulation parameters. The visual component of the work was realised as a single performance, with the output being stored as a series of images. The audio component of the composition was generated using an implementation of the *munger1~* external for *Max/MSP* (Bukvic *et al.* 2007), with the *Open Sound Control* protocol enabling communication between development environments (Wright and Freed, 1997; Wright *et al.* 2003; Wright, 2005). Source audio comprised a range of melodic and noise-based flute samples, from which multiple audio textures were generated using parameters derived from the meso-level analysis of the visual output. In addition, numerous unmapped audio textures were generated independently using static synthesis parameters. To enable the evaluation of the perceptual results of the adopted mapping system, the process of media organisation sought to enhance correspondences between medium morphologies, with perceptual correlations between aural and visual densities informing the macro-level organisation and editing of mapped and unmapped audio textures.

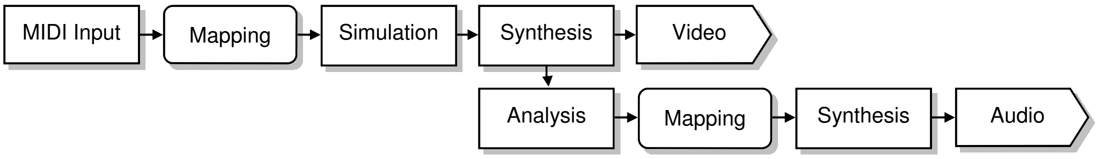

Figure 7.2.1 – A flowchart illustrating the mapping system employed within the creation of *Io*.

The employed mapping system incorporated two independent mapping layers: simulation to visual synthesis and visual analysis to audio synthesis (Figure 7.2.1). Mappings between simulation and visual synthesis parameters were explicit and one-to-one in nature. With the exception of artistic decisions regarding the visual appearance of particles, the products of attractive and repulsive forces upon each particle were rendered directly. The mapping layer between visual analysis and audio synthesis parameters enabled the transposition of visual morphology to the aural domain as a mechanism for the formation of temporal congruence. Explicit mappings within this layer are listed in Table 7.2.1 and were one-to-one and static in nature with value scaling and ramping where required. In addition to the explicit parameter mappings used throughout the process of media generation, an implicit mapping layer between the perceived properties of aural and visual macro-structure was employed to inform the editing and organisation of audio textures. Dynamic mappings within this layer were based on perceptual correspondences between aural and visual densities.

Input Audio File	Mapping			Output Audio File
Trill3.wav	Particle count	→	Channel 1 voices	Texture1.wav
	Average particle brightness	→	Channel 2 amplitude	
	Average particle speed	→	Channel 2 pitch variation	
	Average particle-emitter radius	→	Channel 2 granular voices	
	Frame accumulation duration	→	Channel 1 amplitude	
F5.wav	Particle count	→	Channel 1 grain rate	Texture2.wav
	Average particle brightness	→	Channel 1 voices	
	Average particle speed	→	Channel 2 pitch variation	
	Average particle-emitter radius	→	Channel 2 voices	
	Frame accumulation duration	→	Channel 2 grain length	
Trill11.wav	Particle count	→	Channel 1 voices	Texture3a.wav
	Average particle brightness	→	Channel 1 pitch variation	
Trill9.wav	Average particle-emitter radius	→	Channel 1 grain length	Texture3b.wav
	Frame accumulation duration	→	Channel 1 amplitude	
GLoop103_env.wav	Particle count	→	Channel 1 voices	Texture4a.wav
	Average particle-emitter radius	→	Channel 1 amplitude	
GLoop103_env.wav	Frame accumulation duration	→	Channel 1 grain length	Texture4b.wav

Table 7.2.1 – A table listing the one-to-one parameter mappings employed within the creation of *Io*.

Upon retrospective observation it is apparent that the use of a single particle emission point fundamentally hindered the success of *Io*. While the trajectory curves generated by the software are initially interesting, they rapidly become predictable and the composition appears to devolve into a series of macro-level swells in audiovisual intensity. The perceptual mappings employed throughout the media organisation process are generally more transparent than the strict mappings between visual analysis and audio synthesis parameters. Similarly the cumulative effects of multiple meso-level mappings are often perceived as a macro-level correlation between mediums. An example of this is the correspondence between visual frame accumulation duration and perceptual audio density, in which multiple low-level mappings are perceived as a single one-to-one mapping between visual and aural density.

Additionally, the transposition of temporal structure first from simulation to visual synthesis and then from visual output to audio synthesis reduces the apparent dynamism in structural relationships. Indeed, the composition proves most successful when elements of temporal incongruence are incorporated. A notable example of this is at 1:41, at which the pre-existing relationship between visual and aural density is modified, resulting in a more limited audio spectrum relative to visual density than would be expected. This observation reinforces the theory of temporal consonance and dissonance as presented in Chapter 4. The inclusion of unmapped audio textures also negates some of the predictability throughout the composition, the morphologies of which appear relatively incongruent in comparison to the explicitly mapped components.

7.3 RHEA

The primary aim of *Rhea* was to expand upon the behavioural simulation and parameter mapping systems established throughout the creation of *Io*, and to reduce the inherent predictability of the approach. As such, orbital simulation was used to encourage greater complexity in the particle behaviour. Five randomly selected orbiting masses were defined as particle emission points and imparted attractive forces upon nearby particles. All other orbital components emitted a repulsive force relative to their mass. Location and mass values for each entity were randomly assigned upon creation and remained constant throughout the duration of the composition. Composer interaction was limited to modifications to the particle emission probability, force magnitude scalar and particle colour variables. This reduction in composer influence coupled with the introduction of stochastic behaviours resulted in a more generative system than that used in the creation of *Io*. As with the previous composition, the behavioural simulation and visual synthesis systems were developed in *Processing*, whilst the audio component was generated in *Max/MSP* using a selection of noise-based and melodic samples. Controlling parameters for audio synthesis were derived from the analysis of the visual component and values extracted from the behavioural simulation. As with *Io*, explicitly mapped audio textures were complemented by unmapped material and the organisation of generated media was informed by perceptual correspondences between aural and visual densities.

The mapping system used throughout the creation of *Rhea* incorporated three layers organised into both serial and parallel hierarchies (Figure 7.3.1). Parameters derived from the orbital simulation were explicitly mapped to both visual and audio synthesis systems, while additional control of the granualisation process was derived from emitter-specific data extracted during the rendering process. Explicit mappings used within the composition were all one-to-one and static in nature and were selected to facilitate the creation of a perceptual correspondence between medium densities. All variables extracted from the behavioural simulation and visual analysis systems were subject to value scaling and ramping as required. As with *Io*, a final implicit mapping layer was employed to inform the process of media organisation. All mappings within this layer were derived from perceived correspondences between the apparent densities of each medium.

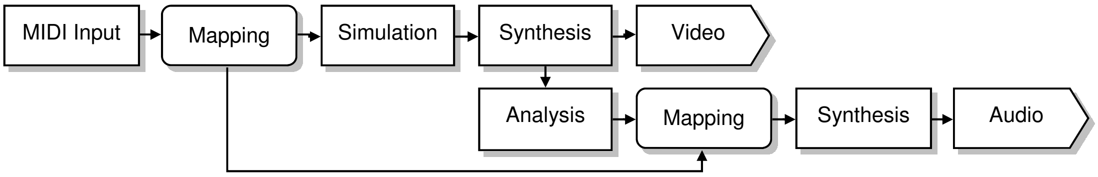

Figure 7.3.1 – A flowchart illustrating the mapping system employed within the creation of *Rhea*.

Unlike the simple transparent mappings used within its predecessor, the mapping system employed in the creation of *Rhea* made use of multiple opaque mappings between simulation parameters and the granular synthesis algorithm (Table 7.3.1). As such, audio derived from simulation data may be perceived as unmapped despite the core mapping being fundamental to the composition process. This highlights the distinction between mappings used to create a perceptual correspondence between audiovisual properties and those used to facilitate the design and composition of a multimedia work. The transparency of the audiovisual mapping system within *Io* arguably contributes to the inherent predictability of the composition, as correspondences between parameters rapidly become uninteresting after being identified by the audience. In contrast, the increased complexity and randomised behaviour of the underlying simulation within *Rhea* generates greater dynamism within the audiovisual morphology, resulting in a less predictable composition. As such, this research finds that while stochastic and generative behaviours can be difficult to control from a compositional perspective, the results will be artistically stimulating due to their inherent unpredictability.

As with *Io*, the primary morphological correspondence throughout *Rhea* is a correlation between the effects of frame accumulation and the density of the audio spectrum, clearly perceivable as a series of audiovisual ‘swells’ that increase in intensity throughout the duration of the composition. The inclusion of unmapped material counteracts some of the predictability prevalent within explicitly mapped media, most clearly demonstrated by the introduction of an unmapped low frequency texture at 1:22. In contrast, the inclusion of granular ‘flourishes’ – brief durations of rapid, synchronised particle emission in both mediums – encourages cross-modal correspondences that maintain aesthetic value despite being largely predictable. On this basis, it would appear that the contrast between elements of temporal consonance and dissonance not only serves to maintain composition interest and reduce predictability, but also enhances moments of strong temporal congruence. Further to this, the divergence and convergence of temporal structures appears to permit the creation of tension and release analogous to the resolution of dissonance within tonal music (see Chapter 4).

Input Audio File	Mapping			Output Audio File
FltArp3.wav	E_0 average particle-centre distance	→	Channel 2 pitch variation	ChordTexture1f.wav
	E_0 average particle-emitter distance	→	Channel 1 amplitude	
	E_0 maximum particle-emitter distance	→	Channel 2 amplitude	
GtrArp3fx.wav	E_0 particle count	→	Channel 1 voices	ChordTexture1g.wav
	Frame accumulation duration	→	Channel 2 voices	
FltArp5.wav	E_1 average particle-centre distance	→	Channel 1 amplitude	ChordTexture2f.wav
	E_1 emitter-centre distance	→	Channel 2 voices	
	E_1 average particle brightness	→	Channel 1 voices	
	E_1 average particle-emitter distance	→	Channel 1 grain length	
GtrArp5fx.wav	E_1 maximum particle-emitter distance	→	Channel 2 grain length	ChordTexture2g.wav
	E_1 particle count	→	Channel 2 pitch variation	
	Frame accumulation duration	→	Channel 2 amplitude	
FltArp6.wav	E_2 average particle-centre distance	→	Channel 1 voices	ChordTexture3f.wav
	E_2 emitter-centre distance	→	Channel 2 grain length	
	E_2 average particle-emitter distance	→	Channel 1 amplitude	
GtrArp6fx.wav	E_2 maximum particle-emitter distance	→	Channel 2 amplitude	ChordTexture3g.wav
	E_2 particle count	→	Channel 2 voices	
	Frame accumulation duration	→	Channel 2 pitch variation	
FltArp7.wav	E_3 average particle-centre distance	→	Channel 2 pitch variation	ChordTexture4f.wav
	E_3 emitter-centre distance	→	Channel 1 pitch variation	
GtrArp7fx.wav	E_3 average particle-emitter distance	→	Channel 1 amplitude	ChordTexture4g.wav
	E_3 particle count	→	Channel 2 amplitude	
	Frame accumulation duration	→	Channel 2 grain length	
BNote1.wav	E_4 average particle-centre distance	→	Channel 1 voices	NoiseTexture1f.wav
	E_4 emitter-centre distance	→	Channel 2 grain length	
	E_4 average particle-emitter distance	→	Channel 1 amplitude	
	E_4 maximum particle-emitter distance	→	Channel 2 amplitude	
	E_4 particle count	→	Channel 2 voices	
	Frame accumulation duration	→	Channel 2 pitch variation	
FltArp2.wav	Average orbiter speed	→	Channel 2 pitch variation	DroneTexture1f.wav
GtrArp2fx.wav	Average orbiter-centre distance	→	Channel 1 amplitude	DroneTexture1g.wav
	Maximum orbiter-centre distance	→	Channel 2 amplitude	
FltArp4.wav	Average orbiter speed	→	Channel 2 amplitude	SparseTexture1f.wav
GtrArp4fx.wav	Average orbiter-centre distance	→	Channel 2 grain rate variation	SparseTexture1g.wav
	Maximum orbiter-centre distance	→	Channel 2 grain length	

Table 7.3.1 – A table listing the one-to-one parameter mappings employed within the creation of *Rhea*. Particle emitters are denoted by E_n.

7.4 DIONE

The primary aim throughout the creation of *Dione* was the analysis of an asynchronous approach to media generation relative to the explicit mapping strategies used within the previous studies. The visual synthesis system employed a set of orbiting elements with random mass and location, around which numerous flocking entities were scattered. Flocking elements were bound by the rules of separation, alignment and cohesion as defined by Reynolds (1987). Each orbiting mass within the simulation exerted a repulsive force upon flocking elements, whilst five randomly selected masses were defined as particle emission points. Emitted particles were subject to an attractive force from flocking elements and a repulsive force from orbiting masses and it is from the interaction of these forces that the resultant visual behaviour was formed. Particle creation probability, force magnitude and particle colour variables were controlled in real-time by the composer, while Brownian deviation was also incorporated within the particle behaviour simulation. As with previous compositions, the visual system was developed in *Processing*.

In contrast to the working processes of both *Io* and *Rhea*, the granular audio textures within *Dione* were developed independently from the visual assets, using granular synthesis parameters based on aesthetic direction rather than an explicit parameter mapping hierarchy. As such, the composition methodology moved away from the proposed framework, exploring the implicit transposition of morphologies from the visual to aural medium through composer derived correlations. This decision was motivated by a desire to evaluate the relevance of implicit parameter mapping techniques within audiovisual composition, relative to the explicit mapping systems employed within previous works. As with previous compositions, the organisation of generated media was based on perceptual correlations between aural and visual densities.

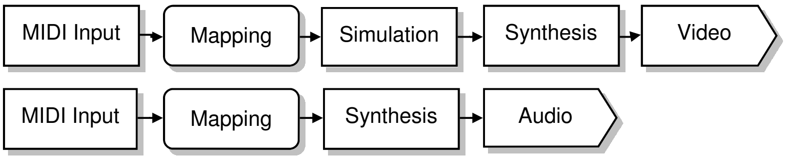

Figure 7.4.1 – A flowchart illustrating the mapping system employed within the creation of *Dione*.

Upon retrospective observation, it would appear that use of solely implicit mappings within the creation of *Dione* has a negative effect on the structure and coherency of the composition. While generalised correspondences between temporal structures are apparent, the dynamic range of temporal congruence is reduced relative to *Rhea*. Similarly, the microstructures of the generated audio textures exhibit reduced dynamism in comparison to those generated within the previous compositions, most likely as a result of the static granular synthesis parameters used within the process of media generation. Based on theory discussed in Chapter 4, such observations are unsurprising; within this particular methodology, temporal congruence can only be formed throughout the process of media organisation, with generated audio bearing no morphological similarity to the visual media, unless specifically imparted by the composer. While this permits the composer greater freedom and control over medium structures, in the case of *Dione* the lack of structural correlations between lower-level morphologies results in a less dynamic composition. It should be noted, however, that such problems could be alleviated by working within a lower-level time scale during the media organisation process, although such an approach would prove significantly more labour-intensive.

7.5 DISCUSSION

In comparing *Rhea* and *Dione*, this research finds that the use of explicit mappings between synthesis algorithms permits the creation of audio material with greater spectral dynamism than that which is generated from static synthesis parameters. Such an observation is unsurprising as the introduction of variant parameters within the process of granular synthesis will typically encourage greater spectral dynamism within the resultant audio output. While low-level morphologies could also be generated through the process of micro-montage or through composer modulation of synthesis parameters, such alternative methodologies move away from the cross-modal analogy from which the audiovisual particles framework is derived.

Within both *Io* and *Rhea* it would appear that the use of explicit mappings typically encourages strong temporal congruence between generated audio gestures and the visual structures from which they are derived. Within both compositions, however, the use of uniformly transparent correspondences between medium morphologies often results in a predictable relationship between medium events. As an alternative approach, temporally dissonant structures derived from opaque or complex mappings appear to offer greater significance to the more congruent elements of the works. On this basis, the presented works illustrate the importance of a dynamic relationship between the temporal structures of each medium, encouraging the creation of tension and release through modulation between congruent and incongruent states.

The use of unmapped media assets offers an alternative mechanism for the creation of dynamic relationships between temporal morphologies. This is most clearly illustrated throughout *Rhea*, in which the interplay between consonant mapped and dissonant unmapped structures creates formalist expectation and suspense (see Chapter 4). Indeed, the introduction of unmapped audio textures at 0:56 and 1:15 appears to offer greater significance to the strongly synchronised granular flourishes at 1:11 and 1:18 respectively. Although dynamic mappings were not used throughout the composition of the works, it is probable that such mappings would have a similar effect, negating the onset of predictability by introducing a dynamic relationship between the temporal structures of each medium.

Upon retrospective observation, the media organisation decisions derived from perceptual correlations between medium densities have been largely successful, with strong correspondences forming the recurring motif for each of the works. This is most apparent within the cross-modal density swells of *Io* (0:47; 0:54; 1:10; 1:18; 1:23; 1:30; 1:46). In this case, however, the consistent repetition of the core correspondence becomes predictable despite changes to the lower-level mappings from which audiovisual structure is derived. Indeed, the composition is most successful when the correspondence is subverted at 1:41, allowing the temporal morphologies to diverge in a manner that contradicts the established audience expectation. As such, this research finds that while compositions devoid of temporal congruence may lack a structural basis from which the audience may derive meaning, if such correlations consistently conform to audience expectation then the composition may rapidly become predictable and uninteresting.

7.6 CONCLUSION

In summary, three implementations of the audiovisual particles framework have been presented, offering example methodologies for the processes of material selection, media generation and media organisation. Cross-comparison of the presented works derived from both experiential and retrospective observations has permitted the discussion of implicit and explicit mapping strategy within the process of audiovisual media generation. Further to this, analysis of the presented methodologies for media organisation has permitted the formation of strategy for the creation and dissolution of temporal congruence within audiovisual media. Further discussion regarding the implementation of explicit mapping strategies within the media generation is presented in the following chapter.

8 Mapping Studies

8.1 INTRODUCTION

This chapter documents four practical applications of parameter mapping strategy within the process of audiovisual media generation. The employed behavioural simulation, synthesis algorithms and parameter mapping hierarchy are presented for each study, followed by a discussion of experiential observations regarding the adopted approach. The chapter concludes with the cross-analysis of resultant media to enable the discussion of parameter mapping strategy within the process of media generation and the implementation of such methodology within the practice of audiovisual composition. All generated media is available on the submitted data disc (see Appendix I).

8.2 MAPPING STUDY 1

The primary aim of *Mapping Study 1* was the exploration of strict meso-level mappings between visual particles and audio grains to facilitate the formation of per-particle correspondences within generated media. The visual synthesis system employed in the study consists of three particle emitters orbiting about a centralised mass. Each emitter within the system was then assigned numerous force components to which particles adhered. The relative magnitudes of each attached force and the probability of particle creation were modified by the composer, permitting abstracted control over the creation and behaviour of particles within the system. The resultant particle motion was then rendered as a series of still images. Audio and video media were generated concurrently from the simulation data in a parallel mapping hierarchy. The audio in *Mapping Study 1* was generated using a custom granular synthesiser built in *Max/MSP* using the *Granular Toolkit* set of externals by Nathan Wolek (2002). The synthesis engine provided real-time control over the amplitude and spatialisation of each grain throughout its duration. Grain creation time and duration were synchronised with that of the particle simulation, while the spatialisation and amplitude of each particle was derived from the location of each particle relative to the centre of the orbital simulation. As such, temporal synchronisation and spatial equivalence form the primary cross-modal correspondences of the study. Due to the strict temporal correlation between mediums the resultant audio grain length was larger than that typically used for granular synthesis. As a result, the behavioural simulation algorithm had no direct control over the micro-structure of the generated granular textures. To ensure that the output audio conformed to the desired aesthetic identity, the employed source media was the product of a custom, unmapped granular synthesiser. As such, while the organisation of meso-level objects within the macrostructure of the audio component is the result of strict mappings between particle behaviour and audio synthesis parameters, the microstructure of generated material should be considered to be unmapped.

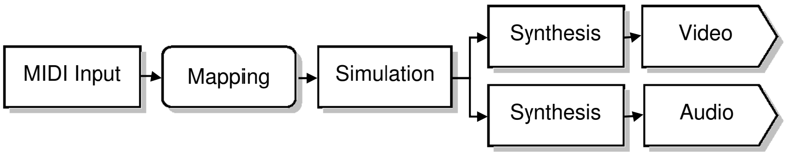

Figure 8.2.1 – A flowchart illustrating the mapping system employed within the creation of *Mapping Study 1*.

The mapping system used in the creation of *Mapping Study 1* (Figure 8.2.1) employed two independent mapping layers: simulation to visual synthesis and simulation to audio synthesis, with the parallel mapping of simulation data to synthesis parameters forming the primary influence upon the resultant media. The mapping process between simulation and visual synthesis required no explicit mapping function as the simulation was rendered directly. Meso-level mappings between simulation and audio synthesis, however, required additional mapping functions to convert the particle location vector to a series of spatialisation parameters relative to the centralised listener location. In addition to the explicit parallel mapping system used in the creation of audiovisual assets, implicit mappings between perceived medium intensities were employed to inform minor edits to the amplitude envelope of the audio data. Although such edits were not originally intended, there existed an apparent disconnect between the intensities of each medium, requiring a reduction in the dynamic range of the audio material to better match that of the visuals. It is probable that this disparity resulted from the centralised listener location and could possibly have been reduced by matching the camera and audio listener locations.

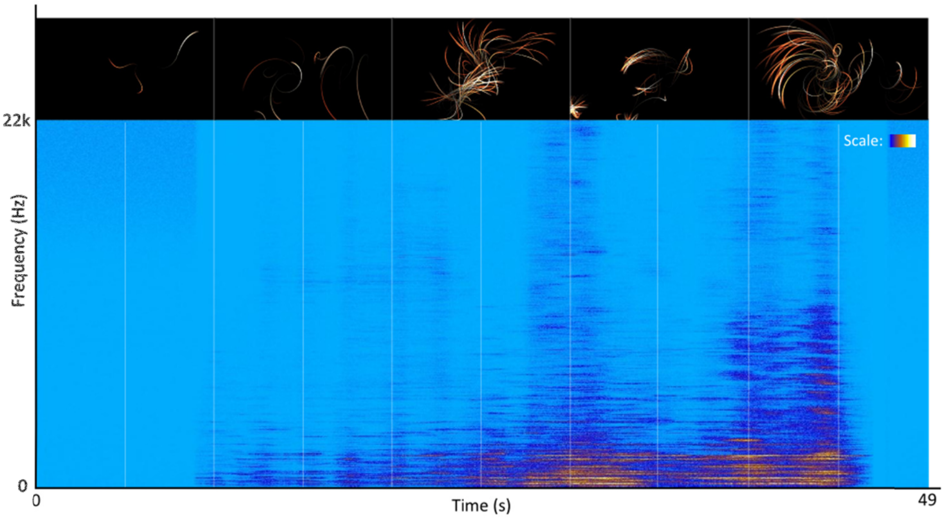

Figure 8.2.2 – A spectrogram illustrating the spectral content of the final audio component for *Mapping Study 1*. The frequency scale is linear.

The mapping process employed in the creation of *Mapping Study 1* enabled the creation of strong temporal congruence between the morphologies of each medium. These correspondences appear to be primarily due to the strict meso-level correlations between particle densities resulting from the parallel creation of particles within each medium. As such, this research finds that the employed meso-level mapping system offers a valid method for the creation of cross-modal correspondences within generated media. Whilst the morphology of the audio meso-structure is largely successful, it is notable that the micro-structure does not offer an equivalent dynamism. Indeed, although modulations within the spectral density of the work are generally successful, the generated audio lacks spectral interest and rapidly becomes predictable. It is probable that this weakness results primarily from the absence of micro-level mappings, indicating a potential limitation of the media generation process.

8.3 MAPPING STUDY 2

The primary aim of *Mapping Study 2* was to examine the effects of alternative mapping approaches within audiovisual media generation through direct comparison with *Mapping Study 1*. The visual synthesis process remained largely unaltered from the previous study, with three particle emitters orbiting about a central mass within three-dimensional space. The probability of particle creation and the magnitudes of various forces could then be controlled by the composer. To ensure consistency, the visual synthesis engine used the same input data as the previous study, with only the randomised emitter starting locations differing. As a result, differences between particle behaviour and the subsequent visual morphology throughout the durations of studies 1 and 2 are purely aesthetic. Whilst the visual synthesis system and the input data remained fundamentally identical to that of its predecessor, the audio synthesis process differed significantly. Audio for *Mapping Study 2* was generated using a custom granular synthesiser built in *Max/MSP*. Unlike the previous study, no direct mapping between independent particles was performed. Instead grain rate, size and pitch were generated randomly by the synthesis algorithm about a composer defined range and midpoint. On this basis, whilst *Mapping Study 1* sought to create strict correlations between the temporal organisation and spatialisation of medium, the methodology for *Mapping Study 2* instead encouraged a more generalised cross-modal correspondence between temporal morphologies.

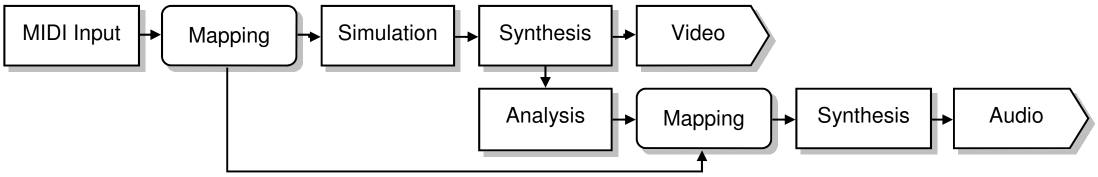

Figure 8.3.1 – A flowchart illustrating the mapping system employed within the creation of *Mapping Study 2*.

The mapping system used within the creation of *Mapping Study 2* (Figure 8.3.1) incorporated three independent mapping layers: simulation to video synthesis, simulation to audio synthesis and visual analysis to audio synthesis. As with *Mapping Study 1* mappings between simulation and visual synthesis required no interpretive functions as the simulation was rendered directly. Mappings between data and audio synthesis parameters employed scalar functions and value ramping where required. All mappings were one-to-one and were selected to ensure the strongest possible correlation between the perceptual densities of each medium (Table 8.3.1). As with the previous study, implicit mappings between the resultant visual output and audio macro-structure were employed to inform the editing of the generated audio. This decision was based on an apparent incongruity between the dynamic ranges of each medium: whilst the perceived intensity of the visual medium exhibited a large dynamic range, the resultant audio texture exhibited far less dynamism in its amplitude envelope. To this end, it was observed that small adjustments to the amplitude envelope of the audio component based on perceived visual intensity noticeably increased the apparent temporal congruence.

Input Audio File	**Mapping**			**Output Audio File**
school_01b_44k24m.wav	E_0 average particle speed	→	Audio buffer scan speed	*ms02a_44k24s.wav*
	E_0 maximum particle-emitter distance	→	Audio buffer scan range	
	E_1 particle count	→	Channel 1 amplitude	
	E_1 maximum particle-emitter distance	→	Channel 1 grain spread	
	E_1 average particle speed	→	Channel 1 grain rate	ms02b_44k24s.wav
	E_2 particle count	→	Channel 2 grain size	
	E_2 maximum particle-emitter distance	→	Channel 2 grain spread	
	E_2 average particle speed	→	Channel 2 amplitude	

Table 8.3.1 – A table listing the one-to-one parameter mappings employed within the creation of *Mapping Study 2*. Particle emitters are denoted by E_n and the audio file used within the final audiovisual output is italicised.

Acknowledging potential disparities between calculated and perceived density values, fill factor calculations offer a basis for the cross-analysis of medium densities (see Chapter 6). By nature, explicit mappings between particle parameters such as those employed in *Mapping Study 1* result in identical fill factor values for each medium, as the particle density and duration values remain constant across mediums. Conversely, the mapping strategy adopted throughout the creation of *Mapping Study 2* allowed independent particle creation time and duration parameters, potentially resulting in different fill factor values for each medium. On this basis, the comparison of fill factor values and perceptual density correspondences enables a discussion regarding the relevance of such correlations in the formation of temporal congruence. The visual output of *Mapping Study 2* consists primarily of low density clouds with particle durations between 1000 and 4000ms, while the audio component consists of high density clouds with durations between 80 and 200ms. If the average (mean) of both component durations and particle density is used, the fill factor for the visual medium throughout the duration of the study is 2.5 x 2.5 = 6.25. Similarly the fill factor for the audio component of the piece is 100 x 0.14 = 14. While this value implies a significant disparity between the theoretical densities of each medium, perceptual comparison of medium densities indicates a strong correspondence. Acknowledging the numerous modulations to particle intensities throughout the mapping process, it is probable that this perceptual correspondence results from the ratiometric nature of medium densities, the comparison of which results in the perception of temporal congruence. Such an observation is unsurprising considering theory presented in Chapter 4; it is the perception of a correlation between medium morphologies that implies temporal congruence rather than any literal equivalence between parameters.

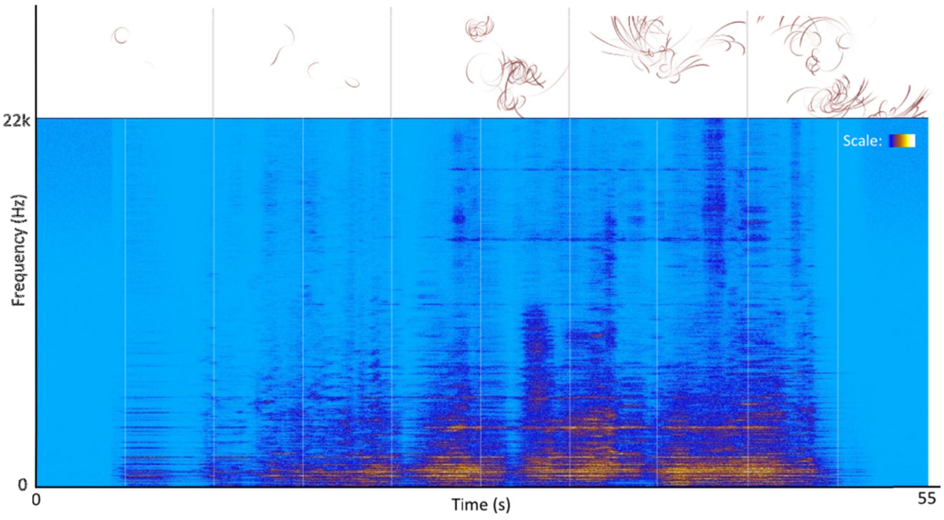

Figure 8.3.2 – A spectrogram illustrating the spectral content of the final audio component for *Mapping Study 2*. The frequency scale is linear.

When comparing the generated visuals with the resultant audio spectrum (Figure 8.3.2), flourishes in spectral activity within the audio appear to be strongly synchronised to increased visual particle activity. This indicates strong correlations between the perceptual densities of each medium. Correspondences between medium intensities are less common, even after macro-level edits to the amplitude envelope of the generated audio. This is unsurprising considering the mapping processes used. It should be noted that while correspondences between medium densities are often clearly perceivable, the mapping process between particle behaviour and audio output may be described as opaque: although an audience may perceive a connection between events in each medium, it is unlikely that the cause of the correspondence will be apparent. This research finds, therefore, that while the adopted mapping strategy enables the generation of temporally congruent media and spectrally diverse audio, it will prove less successful for compositions seeking a clearly interpretable relationship between medium parameters.

8.4 MAPPING STUDY 3

The primary aim of *Mapping Study 3* was to establish the relevance of ‘additive’ visual particle clouds in the formation of temporal congruence. The use of additive blending (Astle and Hawkins, 2004) within visual particle systems permits an increase in visual luminosity relative to a localised increase in particle density, potentially creating temporal congruence through correlations between both medium densities and perceived intensities. The simulation component employed the same orbital particle emission system as the previous studies, with user interaction being limited to modifications to the particle emission probability. Unlike the previous studies, emitted particles had no initial velocity and exhibited only minor alterations to their size and rotation throughout their lifespan. Further to this, rendered particles were significantly larger than in previous works to ensure interesting blending effects. The audio component of the study was generated in the same manner as *Mapping Study 2*, with values derived from both behavioural simulation and visual analysis systems being mapped to the parameters of a custom granular synthesiser built in *Max/MSP*.

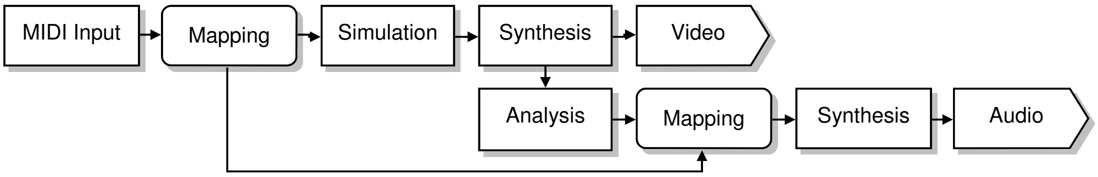

Figure 8.4.1 – A flowchart illustrating the mapping system employed within the creation of *Mapping Study 3*.

As with *Mapping Study 2*, the mapping system employed in the creation of *Mapping Study 3* incorporated two independent mapping layers: simulation to video synthesis and simulation to audio synthesis (Figure 8.4.1). The explicit parameter mappings are listed in Figure 8.4.2, all of which are one-to-one and invariant in nature. As with the previous studies, the resultant visuals were a direct rendering of the simulation data, requiring no interpretive functions. Conversely, value scaling and ramping were employed where required to permit the mapping of data to granular synthesis parameters. The mapping matrix between parameters was chosen to facilitate a perceptual correspondence between cumulative particle formations in each medium, whilst masking the individual mappings from which such correlations may arise. The component mappings within the system can therefore be described as opaque; although the morphologies of each medium are perceived as congruent, the causal parameter mappings from which this relationship is derived are never directly apparent. As with *Mapping Study 2*, generated audio exhibited a limited dynamic range relative to the resultant visual medium and required numerous compensatory edits to the amplitude envelope to improve the perception of temporal congruence. These modifications where informed by implicit mappings between perceived medium properties.

Input Audio File	**Mapping**			**Output Audio File**
school_01b_44k24m.wav	E_0 particle count	→	Audio buffer scan speed	ms03a_44k24s.wav
	E_0 maximum particle-emitter distance	→	Audio buffer scan range	
	E_1 particle count	→	Channel 1 amplitude	*ms03b_44k24s.wav*
	E_1 maximum particle-emitter distance	→	Channel 1 grain length	
	E_2 particle count	→	Channel 2 amplitude	ms03c_44k24s.wav
	E_2 maximum particle-emitter distance	→	Channel 2 grain length	

Table 8.4.1 – A table listing the one-to-one parameter mappings employed within the creation of *Mapping Study 3*. Particle emitters are denoted by E_n and the audio file used within the final audiovisual output is italicised.

Upon retrospective analysis, *Mapping Study 3* exhibits strong perceptual correspondences between medium densities, an observation that is also apparent when comparing the audio spectrum (Figure 8.4.2) to the visual output. This result is not surprising based on observations made throughout the development and analysis of *Mapping Study 2*. Of greater interest is the persistence of these correlations despite significant opacity between input and output parameters within the mapping matrix. The use of additive blending within the visual synthesis system offers the study an interesting aesthetic, with the resultant morphology of visual intensity supplementing the modulations to visual density as perceived in the previous studies. Indeed, the mapping between emitter particle count and audio buffer scan speed appears to create the most significant moments of temporal congruence, with the spectral diversity of the resultant audio corresponding quite clearly to the perceived visual intensity. The use of larger visual particles contributes not only to the resultant visual aesthetic, but also highlights the importance of ratiometric relationships between medium properties in the formation of temporal congruence. Indeed, while studies 2 and 3 have identical particle counts throughout their duration, the increased particle size in the latter implies a greater perceptual density within the visual component relative to the first. Despite this, the perception of temporal congruence is preserved, most likely resulting from a correspondence between the relative morphologies of each medium. As such, this research finds that the process of media generation adopted throughout the creation of *Mapping Study 3* not only offers an interesting aesthetic basis for audiovisual composition, but also preserves the fundamental correspondences between medium morphologies that permit the perception of temporal congruence.

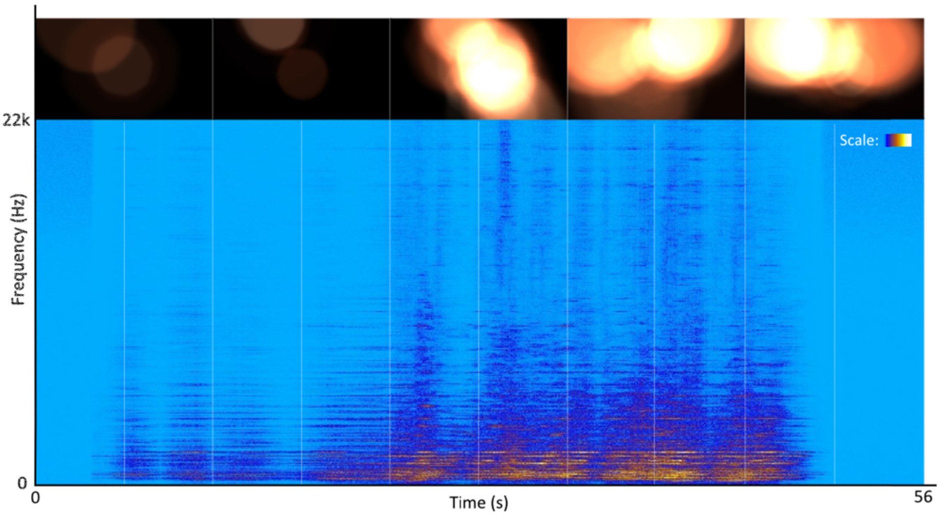

Figure 8.4.2 – A spectrogram illustrating the spectral content of the final audio component for *Mapping Study 3*. The frequency scale is linear.

8.5 MAPPING STUDY 4

Mapping Study 4 returns to the parameter mapping methodology established in *Mapping Study 1*, whilst employing the visual aesthetic of *Mapping Study 3*. Unlike the dynamic meso-level particles used throughout the creation of *Mapping Study 1*, *Mapping Study 4* explored the use of static micro and sound object-level particles with strict synchronisation of particle start time, duration and spatialisation parameters. The primary aim of the study was to establish the effects of strongly synchronised particles in comparison to macro-level correspondences between the temporal morphologies of each medium as observed in studies 2 and 3. The study consists of two component parts utilising the same input data, but with differing particle durations. Study 4a uses synchronised micro-level particles within mediums, whilst study 4b explores the parallel creation of sound-object level particles. By using the same input data, study 4a should, by nature, appear perceptually less dense than study 4b. Unlike *Mapping Study 1*, audio grain size is within the usual range for granular synthesis (Roads, 2004), requiring no pre-granualisation of the source audio material. Further to this, the listener location for audio spatialisation calculations corresponded to the video viewport, resulting in a more predictable audio output that, unlike previous studies, required no compensatory modifications to the amplitude envelope. As such, the spectral content of the final audio output results entirely from the simulation parameters. As with *Mapping Study 3*, video particles were static in nature with large size and additive blending to facilitate perceptual increases in visual intensity. Unlike the dynamic particle updates utilised in *Mapping Study 1*, grain size, duration and location were stored during the visual rendering process. This data was then read by the granular synthesiser to generate audio grains with equivalent amplitude and spatialisation parameters.

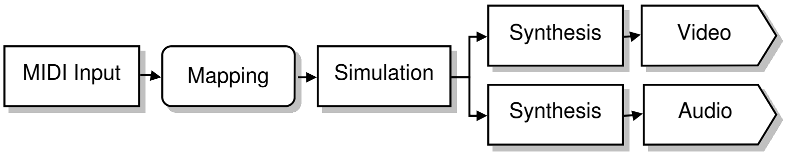

Figure 8.5.1 – A flowchart illustrating the mapping system employed within the creation of *Mapping Study 4*.

The mapping system used throughout the creation of *Mapping Study 4* employed a parallel system of explicit one-to-one mappings that are invariant throughout the both parts of the study (Figure 8.5.1). Mapped values were limited to particle start time, duration and spatialisation. Within the latter, an interpretive function was required to convert the particle location vectors into spatialisation values relative to the listener location. No other value ramping or scalar functions where required. Unlike studies 2 and 3, the generated audio output results purely from the stochastic processes within the simulation; modulations to the particle creation probability and the choice of source audio material are the only points at which the composer may exercise aesthetic control over the resultant audio material. To enable the cross-analysis of generated media, the employed input data and source material was identical for both component studies.

Comparing the audiovisual output and resultant spectrograms (Figure 8.5.2 and Figure 8.5.3) of *Mapping Study 4*, it is apparent that the use of sound-object level particles results in a greater spectral density within the audio domain, and a perceived increase in visual intensity. Such a result is unsurprising; taking an average particle duration of 75ms for study 4a and 350ms for study 4b and assuming a constant grain creation rate between studies, the fill factor when using sound object-level particles is 4.66 times greater than when using micro-level components ($350r \div 75r = 4.66$, where r denotes the grain creation rate). The perceptual effect of this increase in particle density is greater spectral diversity within the audio component, whilst the additive blending of particles results in an increase in perceived visual intensity. Also apparent throughout both component studies is the effects of particle spatialisation, with widely spread particle clouds resulting in a perceptual increase in medium densities relative to those with strongly localised particles.

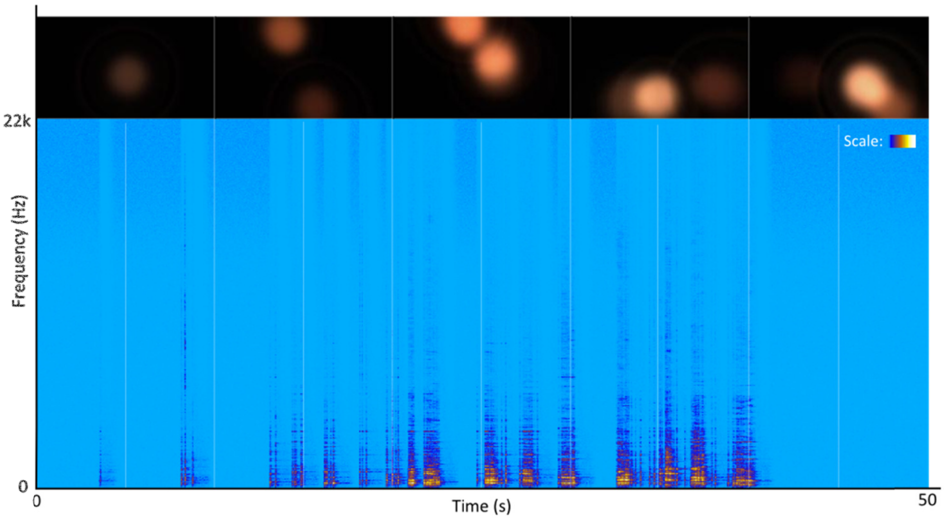

Figure 8.5.2 – A spectrogram illustrating the spectral content of the final audio component for *Mapping Study 4a*. The frequency scale is linear.

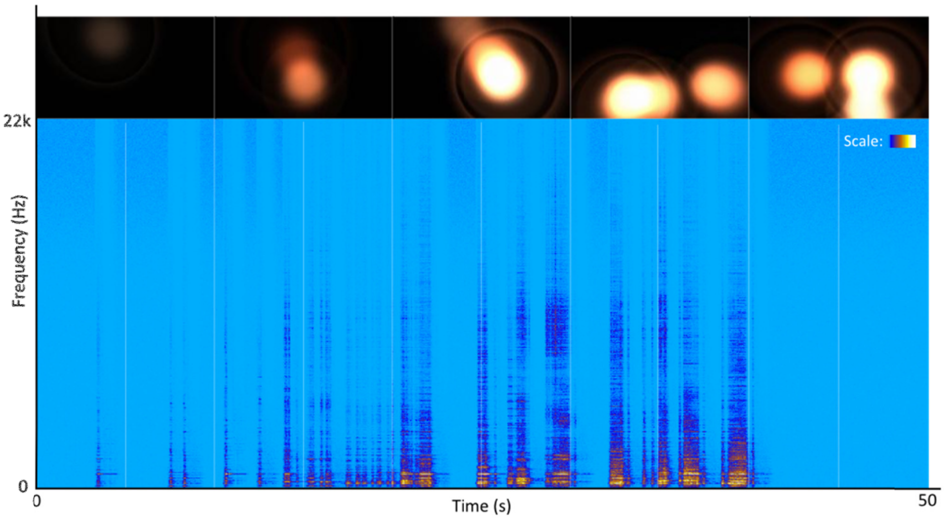

Figure 8.5.3 – A spectrogram illustrating the spectral content of the final audio component for *Mapping Study 4b*. The frequency scale is linear.

The use of strict mappings between particle parameters manifests as a perceivable synchronisation between medium events, the strength of which varies throughout the duration of the component studies. During periods of low particle density, correspondences between individual particles in each medium are clearly apparent. Conversely, during periods of high particle density the relationship between particle parameters becomes more opaque, permitting the perception of generalised correspondences between medium morphologies, but not direct correlations between individual particles. It is interesting to note, however, that the temporal structures of each medium remain largely consonant throughout the duration of the study. As such, this research finds that the mapping strategy adopted throughout the creation of *Mapping Study 4* enables the perception of temporal congruence derived not only from identifiable relationships between individual particles, but also the higher-level analysis of temporal morphology within each medium. It is probable, however, that the strong cross-modal correlations resulting from such an approach may rapidly become predictable if used in isolation, requiring the careful use of such material throughout the media organisation process.

8.6 DISCUSSION

Particle relationships in the presented studies can be divided into two discrete groups. Studies 1 and 4 employ a process of synchronous per-grain synthesis that directly maps the properties of a simulated particle to both visual and aural mediums. This relationship between particles in each medium are defined as *explicit*; ignoring scaling and ramping functions, variables for video and audio synthesis are identical. Conversely, particle relationships in studies 2 and 3 are defined as *implicit*; particles in one medium are inherently affected by those in the other, but there exists no direct relationship between parameters. By definition, particles within an explicit particle relationship will exist within the same organisational level. A simulated micro-level entity would result in a micro-level particle in both visual and aural mediums, for example. Particles within an implicit particle relationship have no such limitation; parameters from one organisational level can be mapped directly to parameters within another, with the internal mechanism of each system controlling the lower-level organisation of the medium. It should be noted that while audiovisual media that exhibit an explicit particle relationship will, by nature, require an explicit mapping system, generated media that employ an implicit particle relationship can be created using any combination of implicit and explicit mappings.

While explicit particle relationships will typically appear more transparent, this is dependent on the perspective from which each medium is viewed. Indeed, it would appear that the perceptual disconnect between medium morphologies in *Mapping Study 1* is due primarily to the different perspectives each medium is perceived from. By using multiple-speaker audio systems or through the use of head-related transfer functions it is possible for a listener to accurately locate audio within three-dimensional space (Toiviainen, 2007). The video medium, however, offers no such flexibility and remains fundamentally two-dimensional with a fixed perspective. The choice of listener location within *Mapping Study 1* emphasises this disparity between perspectives and consequently reduces the apparent temporal congruence of the media. Conversely, by matching the listener and rendering viewport in *Mapping Study 4*, the resultant temporal structures appear significantly more congruent. It should be noted, however, that by imposing the limitations of the visual perspective upon the audio medium, the resultant audio component will, by nature, be less immersive. This research finds, therefore, that while a strict correlation between aural and visual reception perspectives enhances the perceptual synchronisation of events, such an approach may fundamentally limit the success of work.

Analysis of the resultant audio sonograms reveals that studies 2 and 3 exhibit the greatest spectral diversity, whilst studies 1 and 4 exhibit the least. It is probable that the spectral diversity of the resultant audio is due primarily to the source audio used. For each study the source audio was stored in a buffer which is scanned at a user-defined rate. Audio grains are then selected at random from a buffer range defined by the scan position, and a specified window size. On this basis, a low scan speed coupled with a small window size results in relatively low grain diversity, whilst a high scan speed and large window size will result in relatively high grain diversity, assuming spectral variation in the source waveform. In studies 1 and 4 the scan rate and window parameters remain constant throughout the duration of the studies resulting in a grain diversity that is entirely dependent on the spectromorphology of the source material. This can be observed in the audio output of study 4a (0:10), in which the random windowing of a spectrally diverse section of the source waveform results in the most spectrally complex section of the study despite the low particle density during the same period. This potential disparity is alleviated by the mapping processes employed in studies 2 and 3, within which particle parameters are mapped directly to the scan rate and window parameters of the granular synthesis algorithm. The effect of this is most clearly audible in *Mapping Study 3* (0:26, 0:32), in which periods of high particle density result in a relative increase in spectral diversity within the audio. Similar effects can be observed in *Mapping Study 2* (0:37), although the correspondence between morphologies is less apparent. From this, this research concludes that the implicit particle relationships in studies 2 and 3 permit greater control over the modulation of spectral content relative to visual behaviour than the explicit particle relationships employed in studies 1 and 4.

The media generation processes for studies 2 and 3 explore the concept of fill factor as a measure of medium density and offer a conceptual method for preserving correlations between medium densities despite varying organisational time-scales. Whilst analysis of medium fill factor values can inform a perceptual correlation between mediums, the value should only be viewed as a measure of relative density over time. The shortcomings of this terminology become apparent in comparing the fill factor values and perceptual densities of studies 2 and 3. Comparing the visual components of each study a clear disparity between the apparent densities of each medium can be perceived, with *Mapping Study 3* exhibiting a greater perceptual density relative to *Mapping Study 2*. This disparity is due to the different visual rendering approaches employed within each and exists despite the use of identical particle creation probability and duration variables across the studies. It is clear therefore that the apparent density of the visual medium is relative. If the audience perspective of video is defined by the frame in which it exists then the perceptual density of a particle system within this frame will be defined by the ratio of particle activity to inactivity across the screen. A highly localised burst of small particles will, for example, appear less dense than the same number of particles of equivalent duration but greater size spread across the visual frame. The same concept is also true within the audio medium as granular material will have a perceptual density relative to the ‘frame’ in which it is perceived. In comparing the visual component of studies 2 and 3, it is apparent that *Mapping Study 3* exhibits greater visual density, whilst comparisons of the audio output reveal that the perceptual audio density is very similar. When the studies are viewed in isolation, however, this distinction is less obvious. It would appear that if the perceptual medium densities remain ratiometric then a correspondence between temporal structures will be perceptible. Conversely, if modulations in medium density are not relative it is likely that temporal congruence will be reduced regardless of medium fill factor values.

8.7 CONCLUSION

In summary, four approaches to the process of media generation within the audiovisual particles framework have been presented. Cross-comparison of the presented studies informed by experiential and retrospective observations has permitted the analysis of mapping strategies within the context of concurrent audiovisual synthesis. This facilitated the discussion of implicit and explicit particle relationships informed by ratiometric correspondences in medium densities. As such, the presented studies offer example implementations of parameter mapping strategy within the discussed composition framework, leading to the transposition of temporal morphology between mediums and the formation of perceptual congruence between structures.

9 In Perpetuity

9.1 INTRODUCTION

The final practical component of the presented research comprises two linear audiovisual works: *In Perpetuity: The Early Lilacs* and *In Perpetuity: The Linden Trees*. These compositions represent the practical application of formalist approach to audiovisual composition and audiovisual particles framework presented in Chapters 4 and 6 respectively. Unlike the studies presented in Chapters 7 and 8, the primary motivation behind the creation of the *In Perpetuity* compositions was the articulation of composer-led conceptual and aesthetic bases for the creation of audiovisual media. The chapter begins with an overview of the software systems developed to facilitate the creation of the presented work, followed by the definition of a generic methodology for their composition. The implementation of the stated methodology within the context of each composition is then discussed with reference to presented theory. The *In Perpetuity* videos are available in high-definition 720p format on the submitted data disc (see Appendix I).

9.2 SOFTWARE TOOLS

Prior to the composition of the presented works, a range of software tools was developed to facilitate the process of media generation. This includes two *Java* libraries for use within the *Processing* environment and a package of granular synthesis systems developed in *Max/MSP*. The development of each system was informed by theory identified throughout the creation of both the composition studies presented in Chapter 7 and the mapping studies presented in Chapter 8. Further to this, the developed software was subject to modification throughout the creation of the *In Perpetuity* compositions dependent on situational requirements. Within the context of this research, the presented software tools offer a mechanism for the practical application of theory presented in Chapters 4, 5 and 6. While these tools are documented here to inform future practical implementation of the audiovisual particles framework, it should be noted that such implementations are not dependent their use. Indeed, the methodology defined in the audiovisual particles framework may make use of alternative systems, and the tools presented here may be repurposed for use in contexts outside the framework.

9.2.1 PSLib Library

The *PSLib* library is a collection of *Java* classes used for the creation of dynamic particle systems within the *Processing* development environment. While a full discussion of the individual library components is outside the scope of this research, a brief analysis of the core elements is presented here, with the *Java* library, documentation and source code available on the submitted data disc (see Appendix I).

While there is a wide range of literature regarding the implementation of CPU and GPU-based particle system technology (see Chapter 6), there are few open source systems equivalent to the *PSLib* library. Arguably the most established CPU-based particle system library is the *Particle System API* (McAllister, 2000; 2008), a powerful *C++* library for the creation of complex particle effects. The fundamental design of this system (McAllister, 2000) informed the development of the *PSLib* library although the implementation differs greatly. The *JOPS* library (Gomes, 2009) is designed for use with the *Xith3D* (2010) 'scenegraph' engine and offers similar functionality. Currently, the only publicly available particle library for *Processing* is the *PSystem* plugin (Greenwold, 2004), although this deals primarily with the simulation of particle forces rather than the rendering of 'fuzzy' objects as defined by Reeves (1983). As noted in Chapter 6, there are also numerous commercially available software systems that incorporate particle simulation and rendering functionality (Junker, 2006; Hess, 2007; Trenholme and Smith, 2008; Blackman, 2011; Pallamar and Keller, 2011).

While these systems were considered for the creation of the *In Perpetuity* works, the low-level parameter manipulation and data availability afforded by the creation of a custom library outweighed the work required to implement it. Further to this, this research considers that the flexibility afforded by a custom software system permits greater experimentation throughout the composition process, as new software components can be added as required.

The functionality of the library is divided into three interrelated components: *particles*, *behaviours* and *emitters*. The *DynamicParticle* object forms the basis from which all particle objects are derived and stores position, velocity, size, colour and transparency variables. Particles may be assigned a duration of existence, with both size and transparency values being subject to a user-specified morphology throughout this period. It should be noted that within the context of this library, a particle defines an object with no explicit visual representation; while render functions are included for analysis purposes, only the *ImageParticle*, *RibbonParticle* and *TextParticle* objects explicitly address the visual representation of particle information. A *behaviour* defines an iterative modifier to which any derivative of the *DynamicParticle* object can be subject. Further to this, a *force* is defined as a behaviour that exclusively modifies the particle velocity vector. Examples of these include the *MotionBehaviour* class which modifies the particle location based on the stored velocity vector and the *LinearForce* and *ResistiveForce* classes, which simulate directional and resistive forces respectively.

Within the context of the *PSLib* library, an *emitter* is defined as an encapsulating object to which multiple particles can be attached to simplify particle updates and rendering. An emitter can also specify a set of rules to which attached particles must adhere. For example, particles attached to a *TargetEmitter* will automatically tend towards a user-specified set of location vectors stored within the object. An *emission behaviour* is specific to particle emitters and defines an iterative process by which new particles are created and attached to the associated emitter. For example, the *VariantEmission* class facilitates the random creation of particle variants on each update. To achieve recursive emission, all emitter objects within the library are derived from the *DynamicParticle* class and may therefore be attached to other emitters. This functionality permits the creation of complex iterative particle systems with minimal programming.

The library also includes two simulation objects: *BoidParticle* and *OrbiterParticle*. Within the former, the particle velocity vector is modified based on the flocking rules of separation, cohesion and alignment as defined by Reynolds (1987), while the latter incorporates orbital simulation based on the equations stipulated in Newton's law of universal gravitation. These objects are defined as both particles and emitters allowing them to function either as individual entities, or as encapsulating objects for other particles. Further to this, both objects may also be subject to particle or emission behaviours.

9.2.2 PMLib Library

The *PMLib* library is a *Java* implementation of the parameter mapping theory presented in Chapter 5 and facilitates the formation of complex mapping systems within the *Processing* development environment. As with *PSLib*, the *PMLib* library, documentation and source code are available on the submitted data disc (see Appendix I).

While there is significant discourse regarding the implementation of parameter mapping frameworks (see Chapter 5), explicit discussion of software tools that enable the use of such systems remain relatively rare. Bevilacqua *et al.* (2005) present the *MnM Mapping Toolbox*: a set of *Max/MSP* externals that employ matrix operations for the creation of multidimensional linear mappings. Similarly Van Nort and Wanderley (2006) offer the *LoM Mapping Toolbox*, permitting the linear and multi-linear interpolation of data between control space and sound parameter space. Within this framework, *OpenGL* visualisation of the control space is also offered. An alternative approach is presented by Steiner and Henry (2007), who document a series of externals for the creation of complex parameter mapping hierarchies within *Pure Data*. Itself an extension of the *[hid]* toolkit presented by Steiner (2005), the library seeks to encapsulate 'complex mapping algorithms … into software objects, opening up new opportunities for exploration' by allowing mapping to be 'approached as a system of logic' (Steiner and Henry, 2007:2)

While the stated examples enable the creation of complex parameter mapping systems within either *Max/MSP* or *Pure Data*, to the author's knowledge there exists no equivalent software library for use within the *Processing* development environment. On this basis, the *PMLib* library is presented to simplify the design and implementation of parameter mapping systems within this context. The flexible nature of the library allows it to be used within a wide range of contexts outside the field of audiovisual art. Consequently, the library has been released in the public domain to offer a simple and efficient parameter mapping solution for users of the *Processing* development environment.

The functionality of the *PMLib* library can be divided into two core components: a parameter mapping matrix and an event sequencer. The *PMLMatrix* class permits the formation of complex mapping systems and incorporates value scaling and ramping throughout. The system incorporates a tag-based address system that is conceptually similar to that used within the *Open Sound Control* protocol (Wright and Freed, 1997; Wright *et al*. 2003; Wright, 2005). Within this system, input and output parameters are assigned a unique identifier string that can be used to access matrix data. This approach simplifies the use of the library by enforcing a clear and logical data access mechanism and also supports the efficient conversion of matrix data into *Open Sound Control* messages where required.

To enable the recording and playback of mapped data, the *PMLib* library includes the *PMLEvent* and *PMLSequencer* classes. The *PMLEvent* object defines a sequencer *event* and stores an array of floating point values alongside a frame index value and an identifier string. The *PMLSequencer* object provides the core sequencing functionality and permits the storage and playback of *PMLEvent* objects on a per-frame basis. Recorded events can be imported from and exported to text files so that data can be reused within alternative contexts. To simplify data parsing, a *PMLMatrix* object can be specified as the sequencer output with stored event data being routed directly to the specified mapping matrix. Similarly, the *PMLSequencer* object can automatically sample matrix data at a specified frame-interval.

9.2.3 PSL.Granular Package

The *PSL.Granular* package is a collection of granular synthesis systems built using *Max/MSP* and consists of *PSL.GSynth*, *PSL.SGrain* and *PSL.DGrain*: a synchronous/asynchronous granular synthesiser, a static grain cloud generator and a dynamic grain cloud generator respectively. The full package of synthesis systems is available on the submitted data disc (see Appendix I).

Research regarding the practical implementation of granular synthesis techniques typically addresses the design of modular systems for the creation of granular instruments. For example, Wolek (2002) presents a set of *Max/MSP* objects and abstractions that can be employed in the creation of a variety of granular synthesis systems. Similarly, the *GMU* granular synthesis environment presented by Bascou and Pottier (2005) offers a mechanism for the extraction and playback of audio grains. Within each of these systems, the mechanism by which grains are triggered is unspecified, allowing generic operation across a wide range of contexts. Conversely, the *munger1~* external by Bukvic *et al.* (2007) simplifies the implementation of real-time granular synthesis techniques by incorporating a grain triggering system and reducing user control to a series of high-level synthesis parameters.

The primary goal of the *PSL.Granular* package is to simplify the process of grain creation, whilst maintaining real-time control of low-level synthesis parameters with *Open Sound Control* messages. Due to the prevalence of high quality modular synthesis frameworks within the current field, the *PSL.Granular* package is designed as a set of closed instruments rather than a re-invention of pre-existing modular systems. On this basis, the *Granular Toolkit* externals by Nathan Wolek (2002) were employed for the extraction and playback of audio grains, with the development focus instead being the grain triggering mechanism and *Open Sound Control* parameter availability.

The *PSL.GSynth* synthesiser is a custom implementation of granular synthesis capable of both synchronous and asynchronous operation (Roads, 2004). There are two versions of the synthesiser: *PSL.GSynth* and *PSL.GSynthM*. The former incorporates two independent synthesis channels capable of 64 simultaneous grains each, while the latter has one synthesis channel capable of 128 simultaneous grains, permitting greater synchronous grain density. Within each system, audio grains are selected at random from a user specified section of the audio buffer. The position of this window within the buffer can be scanned in real time enabling the granular reconstruction of source material. Synthesis channels include controls for grain creation rate, size, pitch and spatialisation and permit the introduction of random variance on a per-grain basis. Each synthesis parameter is available for modification by *Open Sound Control* messages and incorporates value scaling and ramping to facilitate this.

The *PSL.SGrain* synthesiser is a static grain generator capable of 128 simultaneous grains with independent duration, pitch, spatialisation and amplitude variables. The system offers no internal mechanism for the generation of grain data, being controlled instead by *Open Sound Control* messages or via command sequences created with the *PSLib* and *PMLib* libraries. As with *PSL.GSynth* the synthesis engine incorporates the *Granular Toolkit* externals, with grains being selected at random from a user-specified window within the audio buffer. While conceptually similar to *PSL.SGrain*, the *PSL.DGrain* synthesis system permits the dynamic modification of grain variables during playback via *Open Sound* Control messages or command sequences. Unlike *PSL.SGrain*, the allocation of grain channels within the synthesiser is dependent on control messages, therefore requiring an external dynamic allocation system for the generation of control data. While this makes the use of the synthesiser significantly more complex than *PSL.SGrain*, the dynamic parameter control afforded by such an approach provides a far greater variety of grain cloud evolutions. It should be noted that experimentation with the *PSL.DGrain* synthesis system indicated that the technical complexity of its implementation often outweighed the sonic results. As such, it was not employed in the creation of the *In Perpetuity* series of works. It is included within this thesis as a practical implementation of the parameter mapping methodology identified in the creation of *Mapping Study 1* (see Chapter 8) and as a potential tool for the implementation of the audiovisual particles framework.

9.3 GENERIC METHODOLOGY

Prior to the creation of the presented works a generic composition methodology was defined to inform the processes of material selection, media generation and media organisation. The developed methodology is derived from the audiovisual particles framework presented in Chapter 6 as an example of a morphological approach to audiovisual composition. As such, it defines both the compositional process adopted throughout the creation of the presented works and also the source material from which media was generated. The decision to develop a single methodology for both compositions was based on the desire to create strong conceptual and aesthetic similarities between each work. The proposed methodology is divided into three processes as defined by the audiovisual particles framework: material selection, media generation and media organisation.

9.3.1 Material Selection

As defined in Chapter 6, the process of material selection defines the aesthetic and morphological bases from which audiovisual media are generated. Within the context of the generic composition methodology, this comprised two libraries of source material for use within the media generation systems and source data from which composition morphologies were derived.

The cyclical nature of birth and death formed the thematic inspiration for each of the works, as reflected in the selection of 'In Perpetuity' as the series name. Further to this the titles 'Early Lilacs' and 'The Linden Trees' were adopted from Walt Whitman's *There Was a Child Went Forth* (1855) and Saint Saën's *Danse Macabre* (1874) presenting themes of childhood and death respectively through the detailed description of environmental sights and sounds. This thematic basis is reflected in the selection of source material, with data being derived from UK birth and death statistics for the last century (Beaumont, 2011). Further to this, the selection of audio material was motivated by the relationship between pitch and noise as a metaphoric representation of the transition between life and death, while the aesthetic direction of the visual component was informed by satellite imagery cataloguing the various stages of star formation and expiration.

Based on the adopted conceptual and aesthetic motivation, the audio library consisted of thirteen recordings of crowd noise at a range of public locations. The process of audio acquisition typically involved numerous 'walks' within the selected locations in order to capture elements of pitched conversation amongst the general crowd noise. This method was chosen as a result of observations made throughout the research process, in which transitions between pitched and noise elements within granular source material often offered interesting spectral results following granualisation. Audio material was recorded in 24 bit, 44.1 kHz format and was converted to mono and normalised prior to use within the media generation process. The visual material library comprised eighteen particle texture images and nine colour palettes consisting of one hundred colour values each. Particle texture images were composed from multiple masked Perlin noise layers, while palette values were extracted at random from a selection of colour gradient images, themselves derived from satellite imagery of stars.

9.3.2 Media Generation

The media generation process within the audiovisual particles framework permits the concurrent synthesis of media facilitated by media-driven or data-driven parameter mapping strategies (see Chapter 6). Within the generic methodology, the process of media generation employed four systems: simulation, parameter mapping, visual rendering and audio synthesis. With the exception of audio synthesis, all of the component systems were developed within the *Processing* environment and incorporate the *PSLib* and *PMLib* libraries for particle simulation/rendering and parameter mapping respectively. Audio synthesis was achieved using the *PSL.Granular* synthesis package with data being transferred between development environments using the *Open Sound Control* protocol.

Due to the potential limitations of parameter variance within compositional strategy as observed by Roads (2004) and discussed in Chapter 6, three discrete systems were developed for the media generation process, each employing different control data, simulation system and mapping structure. The simulation systems can be classified as either static or dynamic. In the former, particle structures are randomly generated prior to rendering, the variables of which being subject to modification by composer specified forces or behaviours. Conversely, dynamic simulation systems incorporate randomly generated emitter structures from which multiple particles of finite duration may be emitted. Within a dynamic simulation, both emitter and particle components can be subject to force and behaviour modifiers, enabling the creation of a complex, evolving particle morphology.

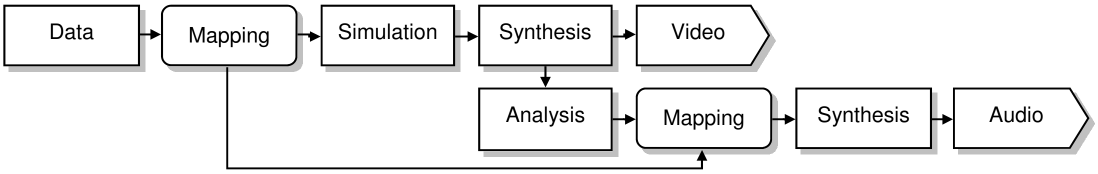

Figure 9.3.1 – A flowchart illustrating the mapping system employed within the creation of the presented works.

The mapping system defined by the generic methodology (Figure 9.3.1) incorporated both data-driven and media-driven components (see Chapter 5) and included three discrete mapping layers: data to simulation, data to audio synthesis and video analysis to audio synthesis. All mappings within each layer were static and either one-to-one or one-to-many in nature, with value ramping and scaling as required. As with the studies presented in Chapters 7 and 8, mappings were derived from perceptual correspondences between medium densities as a mechanism for the formation of temporal congruence.

Each media generation component used source data derived from that specified within the material selection process to modulate variables within simulation and audio synthesis systems. The results of the simulation were then rendered to form the visual output. Informed by the conceptual parallel between the addition of particles in each medium (see Chapter 8) particle textures were additively blended (Astle and Hawkins, 2004) prior to rendering.

To permit the creation of audio material, each simulation component stored two sets of data during the simulation: particle properties and emitter analysis information. The first of these recorded the frame index, duration, intensity and location of each particle generated by the dynamic simulation components. This information was then converted to grain amplitude and spatialisation data for use within the *PSL.SGrain* synthesiser. The second data set stored emitter-specific values, such as average particle speed and average emitter-particle distance, at a composer-specified interval. This was then sent to either the *PSL.GSynth* or *PSL.GSynthM* synthesisers via the video analysis to audio synthesis mapping layer.

While concurrent synthesis was possible with the developed software systems, sequential rendering was adopted so that experimentation with audio source material could be carried out. This ensured the most suitable combination of aural and visual components within the resultant composition. Indeed, as observed in Chapter 6, an iterative approach to media generation, in which the results of a particular combination of source material and mapping hierarchy are analysed to inform future material selection, arguably permits greater diversity within the resultant output.

9.3.3 Media Organisation

The final stage within the generic methodology is that of media organisation. As defined in Chapter 6, this process involves the organisation of generated audiovisual media according to composer intention. The media organisation process was divided into two phases: video editing and audio composition, with audio macro-structure being derived from the established visual morphology. During the video editing phase, the visual output of each media generation component was presented as a single evolving morphology with no editing or modification with the exception of subtle modulations to the visual intensity envelope. In areas where the generated visual meso-structure lacked sufficient gestural identity, visual elements from alternative composition sections were additively blended. The resultant composition sections were then arranged in sequential order and the media rendered to file.

Within the generic methodology, the process of audio composition was predicated by the re-synchronisation of audio textures with the generated visual media to ensure high temporal congruence prior to editing. The process of media organisation was informed by theory presented in Chapter 6, employing modifications to the structure of audio material to both enhance and reduce the perceptual congruence between medium morphologies. Further to this, perceptual dissonances between the temporal structures of each medium were introduced through the inclusion of unmapped material and the temporal relocation of mapped material. While the employed processes were derived from the formalist approach to audiovisual composition and audiovisual particles framework presented in Chapters 4 and 6 respectively, it should be noted that the primary influencing factor throughout the media organisation process was the creation of artistically successful material. On this basis, the articulation of underlying conceptual and aesthetic bases took precedence.

9.4 IN PERPETUITY: THE EARLY LILACS

In Perpetuity: The Early Lilacs is a linear audiovisual work composed using the presented implementation of the audiovisual particles framework. The composition takes the concept of birth as its motivation, adopting a visual aesthetic derived from the initial stages of star formation and utilising data derived from UK birth statistics. Source data was divided into three stages, each 6000 frames in length (Figure 9.4.1), with each section being assigned to a different software system as specified by the generic methodology. To provide the desired audiovisual morphology within the final section, source data was supplemented by an alternative set extracted from composer performance (denoted by CV_1 in Figure 9.4.1). Source material for the media generation process was selected from the developed media libraries through a process of experimentation informed by the aesthetic qualities of the generated material.

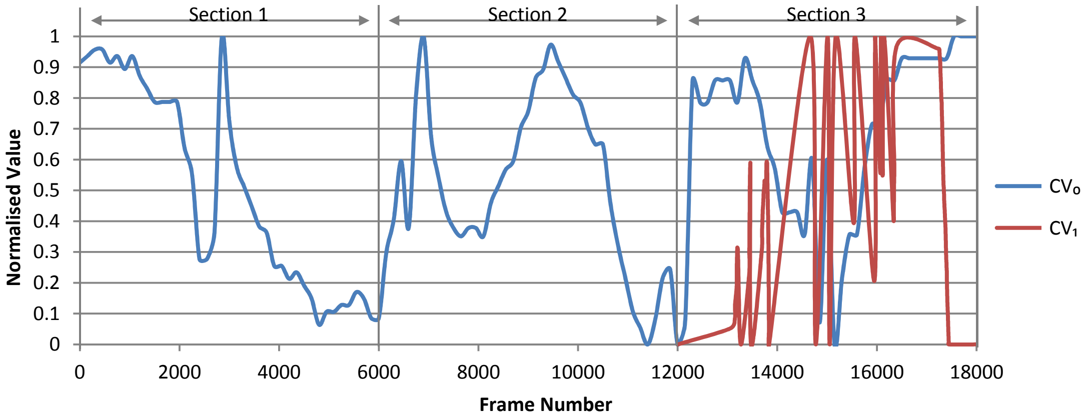

Figure 9.4.1 – A chart illustrating the source data employed within the creation of *In Perpetuity: The Early Lilacs*. The normalised value for each control data series (denoted by CV_n) is plotted against frame number.

Each composition section employed both static and dynamic simulation systems. In the former, flocking behaviour was applied to static particle clouds to encourage subtle modulations in the particle structure. The dynamic simulation component incorporated a centralised particle emitter, with the underlying emission variables being subject to variation between composition sections. The parameter mapping hierarchy adopted throughout the media generation process conformed to that specified within the generic methodology (Figure 9.3.1). Within this, mappings between source data and simulation were limited to particle intensity and duration, emission probability and radius, and boid separation force within the flocking simulation (Reynolds, 1987). Mappings employed throughout the process of audio texture generation are listed in Table 9.4.1 and were all one-to-one or one-to-many and were static in nature (see Chapter 5). Mappings between emitter analysis and audio synthesis variables were derived from perceptual correspondences between particle cloud density and spectral density within the resultant audio texture. To this end, target synthesis variables were limited to audio buffer scan speed and range, and grain size and pitch variance. This decision was informed by observations made throughout the creation of *Mapping Study 3* (see Chapter 8), in which such mappings were observed create a strong temporal congruence between the morphologies of the resultant media.

Source Audio File	Mapping			Output Audio File
school_1d_44k24m.wav	E_0 particle count E_0 average particle-emitter distance	→	Audio buffer scan speed Audio buffer scan range	fs01a_e1_44k24s.wav
school_1c_44k24m.wav	E_0 particle count E_0 average particle-emitter distance	→	Audio buffer scan speed Audio buffer scan range	fs01a_e2_44k24s.wav
city_02_44k24m.wav	E_0 particle count E_0 average particle-emitter distance	→	Audio buffer scan speed Audio buffer scan range	fs01a_e3_44k24s.wav
city_02_44k24m.wav	E_0 particle count E_0 average particle-emitter distance	→	Audio buffer scan speed Audio buffer scan range	fs01a_e4_44k24s.wav
city_02_44k24m.wav	E_0 particle count E_0 maximum particle-emitter distance	→	Audio buffer scan speed Audio buffer scan range	fs01b_e1_44k24s.wav
city_02_44k24m.wav	E_0 particle count E_0 maximum particle-emitter distance	→	Grain size Grain pitch variance	fs01b_e2_44k24s.wav
water_01_44k24m.wav	E_0 particle count E_0 maximum particle-emitter distance	→	Audio buffer scan speed Audio buffer scan range	fs01b_e3_44k24s.wav
school_1e_44k24m.wav	E_0 particle count E_0 maximum particle-emitter distance	→	Audio buffer scan speed Audio buffer scan range	fs01b_e4_44k24s.wav
city_01_44k24m.wav	CV_0 data CV_0 data	→	Audio buffer scan speed Audio buffer scan range	fs01b_d1_44k24s.wav
city_01_44k24m.wav	CV_0 data CV_0 data	→	Audio buffer scan speed Audio buffer scan range	fs01b_d2_44k24s.wav
city_01_44k24m.wav	CV_0 data CV_0 data	→	Audio buffer scan speed Audio buffer scan range	fs01b_d3_44k24s.wav
school_1a_44k24m.wav	CV_0 data CV_0 data	→	Audio buffer scan speed Audio buffer scan range	fs01b_d4_44k24s.wav
school_1a_44k24m.wav	CV_0 data CV_0 data	→	Audio buffer scan speed Audio buffer scan range	fs01b_d5_44k24s.wav
school_1b_44k24m.wav	CV_0 data CV_1 data	→	Audio buffer scan range Audio buffer scan speed	fs01c_d1_44k24s.wav
school_1b_44k24m.wav	CV_0 data CV_1 data	→	Audio buffer scan range Audio buffer scan speed	fs01c_d2_44k24s.wav
school_1b_44k24m.wav	CV_0 data CV_1 data	→	Audio buffer scan speed Grain pitch variance	fs01c_d3_44k24s.wav

Table 9.4.1 – A table listing the parameter mappings employed within the creation of In Perpetuity: *The Early Lilacs*. Particle emitters are denoted by E_n and control data is denoted by CV_n.

Relationships between source data and audio synthesis parameters were derived from experimentation with the aesthetic and spectral qualities of the resultant media. As such, mappings were limited to audio buffer scan speed and range, and grain pitch (Table 9.4.1). In addition to media generated through implicit particle relationships (see Chapter 8), audio material was also generated using explicit particle relationships derived from simulated particle data. Following conversion from the particle data generated by the simulation component to grain amplitude and spatialisation values, numerous audio textures where generated, each exhibiting strict correlations between particle information within each domain and, as such, a high level of temporal congruence. To ensure the desired spectral aesthetic within the resultant media, the majority of the source audio employed throughout this process was pre-granulised (see Chapter 8). All generated audio material used in the creation of *In Perpetuity: The Early Lilacs* is listed in Appendix II and is included on the submitted data disc.

The media organisation process adopted throughout the creation of *In Perpetuity: The Early Lilacs* conformed to that defined within the generic composition methodology, involving the manipulation of audio morphologies to modulate temporal congruence throughout the resultant macro-structure. As stated in the generic methodology, the primary motivation throughout the process of media organisation was the articulation of the adopted conceptual and aesthetic bases rather than an explicit study of temporal congruence within audiovisual media.

9.5 IN PERPETUITY: THE LINDEN TREES

In Perpetuity: The Linden Trees adopts the concept of death as artistic motivation, employing data derived from UK death statistics for the last century and a visual aesthetic derived from satellite imagery cataloguing the various stages during the expiration of sun-like stars. As with *In Perpetuity: The Early Lilacs*, source data was divided into three sections (Figure 9.5.1), with each assigned to a different software simulation as specified by the generic composition methodology. To produce a more complex, ‘polyphonic’ audiovisual morphology and to encourage gestural similarities between the two compositions, the primary control data set was supplemented by an alternative set composed from sections of the birth data employed within the creation of *In Perpetuity: The Early Lilacs*. These data sets are illustrated in Figure 9.5.1 by CV_0 and CV_1 respectively. Further to this, two additional data sets were derived from both the product and the average (mean) of the primary data values. These are illustrated in Figure 9.5.1 by CV_2 and CV_3 respectively. Source material for both aural and visual components was selected from the media libraries defined within the generic methodology, with employed audio material being listed in Table 9.5.1.

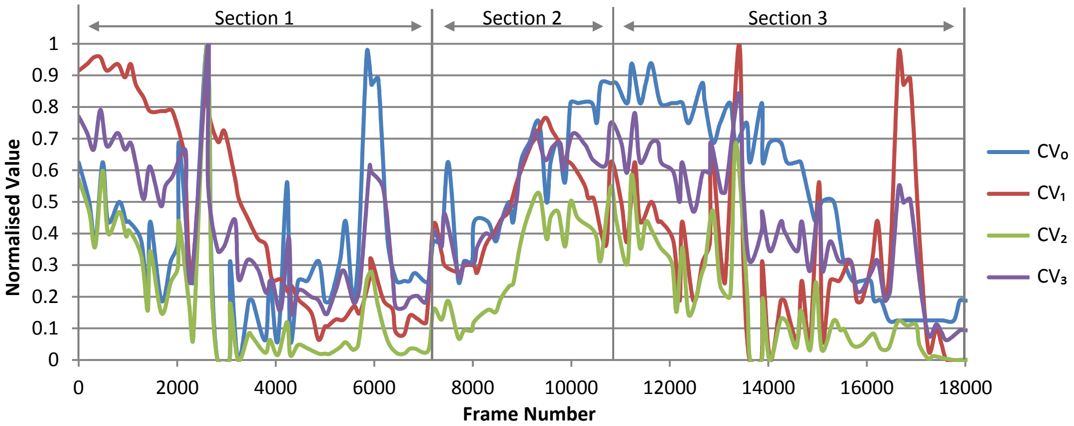

Figure 9.5.1 – A chart illustrating the source data employed within the creation of *In Perpetuity: The Linden Trees*. The normalised value for each control data series (denoted by CV_n) is plotted against frame number.

As with *In Perpetuity: The Early Lilacs*, the simulation systems employed throughout the creation of *In Perpetuity: The Linden Trees* consisted of both static particle structures and complex dynamic emission systems. In the former, flocking behaviour was applied to randomly generated particle clouds to encourage subtle modulations in the resultant particle structure. The dynamic particle system consisted of multiple emitters orbiting about a central mass, each having an attractive force relative to its size. Emitted particles were subject to both the behaviours defined within the parent emitter and also the interaction of attractive forces within the orbital simulation. While the fundamental structure of the dynamic emission system remained constant throughout each composition section, the component variables differed greatly, allowing a wide range of particle morphologies to be generated.

Source Audio File	**Mapping**			**Output Audio File**
park_01e1_44k24m.wav	E_0 particle count E_0 maximum particle-emitter distance	→	Audio buffer scan speed Audio buffer scan range	fs02a_e1_44k24s.wav
park_0e1_44k24m.wav	E_1 particle count E_1 average particle-emitter distance	→	Grain pitch variance Audio buffer scan speed	fs02a_e2_44k24s.wav
schoo1_1b_44k24m.wav	E_2 particle count E_2 maximum particle-emitter distance	→	Audio buffer scan speed Audio buffer scan range	fs02a_e3_44k24s.wav
park_03e2_44k24m.wav	E_0 particle count E_0 maximum particle-emitter distance	→	Audio buffer scan speed Grain pitch variance	fs02b_e1_44k24s.wav
park_04e1_44k24m.wav	E_1 maximum particle-emitter distance E_2 average particle-emitter distance	→	Audio buffer scan speed Grain pitch variance	fs02b_e2_44k24s.wav
school_1c_44k24m.wav	E_3 particle count E_3 maximum particle-emitter distance	→	Grain pitch variance Audio buffer scan range	fs02b_e3_44k24s.wav
park_01e1_44k24m.wav	E_0 particle count E_0 maximum particle-emitter distance	→	Audio buffer scan speed Audio buffer scan range	fs02c_e1_44k24s.wav
school_1e_44k24m.wav	E_1 particle count E_1 maximum particle-emitter distance	→	Audio buffer scan speed Grain pitch variance	fs02c_e2_44k24s.wav
school_1d_44k24m.wav	E_2 particle count E_2 average particle-emitter distance	→	Grain pitch variance Audio buffer scan speed	fs02c_e3_44k24s.wav
park_04e1_44k24m.wav	CV_0 data CV_1 data	→	Audio buffer scan speed Grain pitch variance	fs02a_d1_44k24s.wav
park_03e1_44k24m.wav	CV_0 data CV_1 data	→	Audio buffer scan speed Grain pitch variance	fs02a_d2_44k24s.wav
school_1a_44k24m.wav	CV_0 data CV_1 data	→	Audio buffer scan range Audio buffer scan speed	fs02a_d3_44k24s.wav
park_03e1_44k24m.wav	CV_0 data CV_1 data	→	Audio buffer scan speed Audio buffer scan range	fs02b_d1_44k24s.wav
park_01e2_44k24m.wav	CV_0 data CV_1 data	→	Grain pitch variance Audio buffer scan speed	fs02b_d2_44k24s.wav
school_1b_44k24m.wav	CV_0 data CV_1 data	→	Audio buffer scan speed Audio buffer scan range	fs02b_d3_44k24s.wav
park_01e1_44k24m.wav	CV_0 data CV_1 data	→	Audio buffer scan speed Audio buffer scan range	fs02c_d1_44k24s.wav
park_03e2_44k24m.wav	CV_0 data CV_1 data	→	Grain pitch variance Audio buffer scan speed	fs02c_d2_44k24s.wav
school_1e_44k24m.wav	CV_0 data CV_1 data	→	Audio buffer scan speed Audio buffer scan range	fs02c_d3_44k24s.wav

Table 9.5.1 – A table listing the parameter mappings employed within the creation of *In Perpetuity: The Linden Trees*. Particle emitters are denoted by E_n and control data is denoted by CV_n.

The parameter mapping system adopted throughout the creation of *In Perpetuity: The Linden Trees* conformed to the structure defined within the generic methodology (Figure 9.3.1). As with *In Perpetuity: The Early Lilacs*, mappings between source data and particle simulation systems were limited to particle intensity and duration, and emission probability. Mappings between source and generated data values and audio synthesis parameters are listed in Table 9.5.1 and were all static and one-to-one or one-to-many in nature. As defined by the generic methodology, the conceptual basis for mappings throughout the process of audio texture creation was the creation of perceptual correspondences between medium densities as a mechanism for the creation of temporal congruence. On this basis, target synthesis parameters were limited to audio buffer scan speed and range, and grain pitch variance. As with *In*

Perpetuity: The Early Lilacs, this decision was informed by observations made throughout *Mapping Study 3* (see Chapter 8) and experimentation throughout the media generation process to facilitate the creation of spectrally diverse and aesthetically relevant material. In addition to emitter analysis and source data mappings a selection of audio textures were generated using grain amplitude and spatialisation values derived from the particle simulation data and pre-granulised source audio material. As with *In Perpetuity: The Early Lilacs*, material generated in this manner exhibits a high initial level of temporal congruence due to the strict micro and sound-object level correspondences between the resultant structures. All audio material generated for *In Perpetuity: The Linden Trees* is listed in Appendix III and is available on the submitted data disc.

The media organisation process employed throughout the creation of *In Perpetuity: The Linden Trees* conforms to that specified within the generic composition methodology as a mechanism for the creation of a dynamic level of temporal congruence within the resultant audiovisual macro-structure. As with *In Perpetuity: The Early Lilacs*, the primary focus throughout the process of media organisation was the implementation of conceptual and aesthetic bases as stated within the generic methodology.

9.6 DISCUSSION

Throughout the creation of the *In Perpetuity* works, the audiovisual particles framework provided a strong methodological basis for the creation of audiovisual media and its resultant organisation within composition macrostructure. Implementation of the framework was simple and the adoption of data-driven media generation systems provided a wealth of audiovisual material with a range of temporal structures.

As observed in Chapter 7, the use of stochastic processes potentially limits composer interaction with the resultant media. Indeed, throughout the creation of both the *In Perpetuity* works, many particle structures resulted from stochastic particle emission from the randomised orbital system. Composer interaction with such structures was limited to defining the range from which random values are extracted, the probability of particle emission and the transparency of rendered particles. The inherent unpredictability of the system is both positive and negative; while the use of stochastic or algorithmic processes often results in diverse and interesting micro and sound-object level morphologies, it is often difficult to reproduce these within alternative contexts. For the *In Perpetuity* works, this unpredictability was tempered by using explicit data scripts, but planning for future implementations of the audiovisual particles framework should take such restrictions into account.

From a technological perspective, the software tools functioned as expected, permitting the generation of a wide range of audiovisual material with varying levels of temporal correspondence. The flexibility of the software observed throughout the creation of the *In Perpetuity* works indicates the potential for the repurposing of the systems within alternative implementations of the audiovisual particles framework or entirely different composition methodologies.

The primary concern with the software systems is performance; both the *PSL.Granular* package and the *PSLib* particle system library exhibit limitations with the achievable particle density. For the former, the extra performance overhead of *Max/MSP* greatly reduced the capability of the system. A *Java* or *C++* implementation would significantly improve processing efficiency in this regard. The performance limitations observed within the *PSLib* library primarily result from the reliance on the CPU for particle calculations. A GPU-based particle system would almost certainly yield better performance (Kipfer *et al.*, 2004; Kolb *et al.*, 2004; Latta, 2004), or even a *C++* implementation of a CPU-based system using *OpenGL* (Astle and Hawkins, 2004) rather than the *Java* libraries included with *Processing*. Within the context of this research, the time required to develop such systems would have outweighed the aesthetic benefits and could have potentially limited the data available for parameter mapping. Improved performance within the software systems could, however, prove beneficial for future works that explore greater particle densities.

From an aesthetic perspective, both works exhibit strong thematic correspondences between mediums. Indeed, the additive nature of the particle systems creates a 'celestial' visual aesthetic that correlates thematically with the time-stretched product of the granular synthesis process. Similarly, the adoption of a generic methodology creates an aesthetic consistency between the two works that reinforces the adopted conceptual basis. For example, *In Perpetuity: The Linden Trees* returns almost full circle to the audiovisual aesthetic established at the beginning of *In Perpetuity: The Early Lilacs*, reflecting the cyclical nature of life and death.

Further to this, the slow evolutions in audiovisual morphology and the minimal use of fast cuts in each medium correspond to the conceptual basis for the works; it is the transition between states that is the primary output rather than the states themselves. It would appear, however, that this concept of transitive evolution also limits the aesthetic diversity of the works. Indeed, while multiple particle structures were employed for each composition section and numerous audio textures were generated the audiovisual spectrum of the works remains limited. While this is appropriate within the context of the *In Perpetuity* works, greater diversity within the material selection process and experimentation with alternative macrostructure throughout the process of media organisation could be of benefit to future implementations of the audiovisual particles framework.

The adopted media generation processes enabled the creation of strong temporal congruence at various points within each of the works. The audiovisual output demonstrates a dynamic relationship between temporal structures, with temporally dissonant gestures resolving into passages of monophonic consonance between mediums. It would appear, however, that the resolution of temporal dissonance through the use of 'surges' in audiovisual intensity is slightly overused. As discussed in Chapter 7, this cross-modal correspondence results from an accumulation of low-level correspondences between the temporal structures, which then manifest as a strong meso-level correlation between medium structures. On this basis, future works should employ a greater degree of temporal dissonance, introduced through modification to the adopted mapping structure, or through structural decisions made throughout the media organisation process.

9.7 CONCLUSION

The *In Perpetuity* series of artworks serve to demonstrate the relevance of the audiovisual particles framework. Each work exhibits dynamic, polyphonic relationships between medium morphologies. This is a product of micro and sound-object level transposition of morphology within the media generation process and the meso and macro-level arrangement of generated material within the process of media organisation. The presented works demonstrate the achievement and understanding of all proposed research aims and serve to exemplify the relevance of developed theory within the field of abstract digital animation.

10 Conclusion

10.1 SUMMARY

This thesis has outlined the audiovisual particles framework as a methodological approach to the composition of abstract audiovisual media. Further to this, the practical application of proposed theory has been demonstrated with both the composition and mapping studies and the *In Perpetuity* series of audiovisual artworks. As evidenced by the documented practical components, the audiovisual particles framework presents a valid conceptual and methodological basis for the creation of abstract audiovisual media. To the best of the author's knowledge there is no equivalent methodological framework for composition within the field of abstract digital animation, or the more generalised field of audiovisual art.

Literature regarding the perception and cognition of audiovisual material has been referenced to permit the identification of temporal congruence as a key factor in the manipulation of cross-modal correspondence. The temporal variance of medium parameters has been defined as a temporal morphology and the resultant relationship between medium morphologies has been identified as the primary manipulator of temporal congruence within audiovisual media. Prevalent formalist theory has been evaluated with specific relevance to the interrelationship of medium structures, leading to the definition of a model for audiovisual composition.

To inform the application of presented theory to the process of audiovisual composition, techniques and terminology relevant to the implementation of parameter mapping within audiovisual media generation have been identified. Further to this, data-driven and media-driven mapping hierarchies have been presented as a mechanism for the transposition of temporal morphology between mediums. Following this, a conceptual parallel between audio grains and visual particles has been established to permit the analysis of audiovisual particle structures in terms of density and fill-factor. On this basis, the audiovisual particles framework has been proposed, offering a methodological basis for the composition of abstract audiovisual art with granular synthesis and visual particle systems. Within this, the processes of material selection, media generation and media organisation have been identified to enable the creation and abstraction of temporal congruence within resultant media whilst maintaining full artistic control of the resultant audiovisual aesthetic.

To demonstrate the proposed theory, a series of composition studies has been presented to facilitate analysis of those factors influencing its implementation. Further to this, a series of mapping studies has been documented to aid the discussion of parameter mapping within the media generation process. Derived from posited theory and experiential observations made throughout the creation of the documented studies, *In Perpetuity: The Early Lilacs* and *In Perpetuity: The Linden Trees* have been presented alongside the software systems by which they were created. These compositions represent the culmination of the research process, acting as both demonstrative examples of proposed theory, but also as artworks relevant to the field of abstract digital animation.

The presented research represents the fulfilment of each of the aims stated in the thesis introduction. Perceptual correspondences between aural and visual mediums have been explored, referencing literature from the fields of audiovisual art and cross-modal perception and cognition. A formalist approach to audiovisual composition has been posited as a conceptual basis for the composition of audiovisual media. Parameter mapping discourse has been identified to permit the formulation of the audiovisual particles framework as a practical methodology for the creation of abstract audiovisual art using granular synthesis and visual particle systems. Finally, a number of audiovisual compositions have been presented to demonstrate the practical application of proposed theory.

10.2 FUTURE RESEARCH

As discussed in Chapter 6, the adoption of granular synthesis systems within the audiovisual particles framework is based on conceptual parallels between the organisational structures of audio grains and visual particles. Consequently, audio material generated throughout the documented audiovisual works has made exclusive use of synchronous and asynchronous granular synthesis processes as defined by Roads (2004). While aesthetic differences between the visual components of the composition studies and the *In Perpetuity* series illustrates the potential for a diverse range of visual forms derived from alternative particle rendering systems, it is arguable that the generated audio does not exhibit the same variance. While this consistent sonic aesthetic forms a unifying artistic identity throughout the documented works, it also indicates a direction for future research. The development and implementation of alternative granular synthesis systems, such as those listed in Roads (2004), may allow a more diverse sonic palette whilst maintaining the established conceptual model, thus improving the flexibility of the framework within alternative contexts.

Perhaps the most natural extension to this research is the application of proposed theory within real-time media generation contexts, such as the fields of generative/interactive composition and live cinema/VJ performance. Indeed, the media generation process within the audiovisual particles framework is derived largely from literature within the various fields of interaction and real-time performance. Further to this, the use of parameter mapping as a mechanism for concurrent synthesis of aural and visual media permits a relatively simple translation of software processes. There remains, however, a significant caveat within the potential real-time application of the proposed framework. This thesis has stressed the importance of the media organisation process within the proposed framework to allow artistic control over the resultant audiovisual morphologies and the modulation of temporal congruence within the composition macro-structure. Within the context of real-time performance, however, this mechanism is arguably impractical, thus an alternative system for the organisation of generated media is required. One option would be the removal of the process altogether, with the generated media being output immediately. While such a system would likely require a more complex, dynamic mapping system (see Chapter 5), it would arguably reduce composer intervention, making it best suited to interactive or installation contexts. An alternative approach would be development and implementation of a real-time mechanism for media organisation, permitting composer intervention between media generation and audiovisual output. It is likely, however, that such a system would be very complex, making it better suited to live cinema and VJ performance contexts.

As a practical consideration for the process of media generation, it should be noted that the implementation of the software tools presented in Chapter 9 is dependent on knowledge of object-oriented programming and the *Java* syntax. On this basis, there exists a technological constraint upon the practical implementation of these systems. A potential direction for future research, therefore, would be the development of a unified media-generation environment within *Processing*. Such a system would support the dynamic creation and modification of particle structures as a mechanism to interactively control the resultant audiovisual output. This approach would negate the current reliance of the media generation process upon extensive programming, thus facilitating the adoption of the framework within a wider range of contexts and potentially permitting a stronger composer focus on the artistic intention of the work. Overall, it is felt that the *In Perpetuity* works evidence the potential for the audiovisual particles framework within a variety of artistic contexts and that any developments to further improve the ease of its implementation would be of benefit to the general field of abstract audiovisual art.

References

Abbado, A. (1988) ‘Perceptual Correspondences of Abstract Animation and Synthetic Sound’. *Leonardo Music Journal,* 1 (Electronic art supplemental issue), pp. 3-5.

Alexander, A. & Collins, N. (2008) ‘Live Audiovisuals’. *In:* N. Collins & J. d'Escriván eds. *The Cambridge Companion to Electronic Music*. Cambridge: Cambridge University Press, pp. 126-139.

Alves, B. (2005) ‘Digital Harmony of Sound and Light’. *Computer Music Journal,* 29 (4), pp. 45-54.

Arfib, D., Couturier, J., Kessous, L. & Verfaille, V. (2003) ‘Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces’. *Organised Sound,* 7 (02), pp. 127-144.

Astle, D. & Hawkins, K. (2004) *Beginning OpenGL game programming.* Boston: Thomson Course Technology.

Bailey, N., Fells, N. & Moody, N. (2007) ‘Ashitaka: An Audiovisual Instrument’. *Proceedings of the 2007 international Conference on New Interfaces for Musical Expression.* pp. 148.

Bascou, C. & Pottier, L. (2005) ‘GMU, a Flexible Granular Synthesis Environment in Max/MSP’. *Proceedings of the Sound and Music Computing Conference (SMC’05), Salerno, Italy.*

Battey, B. (2006) ‘Autarkeia Aggregatum: Autonomous Points, Emergent Textures’. *ACM SIGGRAPH 2006 Sketches.* ACM, pp. 92.

Beaumont, J. ed. (2011) *Population.* Social Trends 41 ed. London: The Office for National Statistics.

Bevilacqua, F., Müller, R. & Schnell, N. (2005) ‘MnM: a Max/MSP mapping toolbox’. *Proceedings of the 2005 Conference on New Interfaces for Musical Expression.* Singapore: National University of Singapore, pp. 85.

Blackman, S. (2011) *Beginning 3D Game Development with Unity.* New York: Apress.

Blackwell, T. & Young, M. (2004) ‘Self-organised music’. *Organised Sound,* 9 (02), pp. 123-136.

Bolivar, V.J., Cohen, A.J. & Fentress, J.C. (1994) ‘Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action’. *Psychomusicology: Music, Mind and Brain,* 13 (1-2), pp. 28-59.

Boltz, M.G. (2004) ‘The cognitive processing of film and musical soundtracks’. *Memory & Cognition,* 32 (7), pp. 1194-1205.

Boltz, M.G., Ebendorf, B. & Field, B. (2009) ‘Audiovisual Interactions: The Impact of Visual Information on Music Perception and Memory’. *Music Perception,* 27 (1), pp. 43-59.

Bregman, A. (1994) *Auditory Scene Analysis: The Perceptual Organization of Sound.* 1st ed. Cambridge: MIT Press.

Brougher, K. (2005) ‘Visual-Music Culture’. *In:* K. Brougher & J. Zilzcer eds. *Visual Music: Synaesthesia in Art and Music Since 1900* New York: Thames and Hudson, pp. 88-179.

Bukvic, I.I., Kim, J., Trueman, D. & Grill, T. (2007) ‘munger1~: Towards a Cross-Platform Swiss-Army Knife of Real-Time Granular Synthesis’. *ICMC 2007 Proceedings*. Copenhagen: ICMC.

Cazden, N. (1945) ‘Musical Consonance and Dissonance: A Cultural Criterion’. The Journal of Aesthetics and Art Criticism, 4 (1), pp. 3-11.

Chion, M. (1983) Guide des Objets Sonores. Paris: Éditions Buchet/Chastel.

Chion, M. (1994) Audio-Vision: Sound on Screen. New York: Columbia University Press.

Cohen, A.J. (2005) ‘How music influences the interpretation of film and video: Approaches from experimental psychology’. *Selected Reports in Ethnomusicology: Perspectives in Systematic Musicology,* 12 (1), pp. 15-36.

--- (2002) ‘Music Cognition and the Cognitive Psychology of Film Structure’. *Canadian Psychology,* 43 (4), pp. 215-232.

--- (2001) ‘Music as a Source of Emotion in Film’. *In:* P. Juslin & J. Sloboda eds. *Music and Emotion: Theory and Research*. Oxford: Oxford University Press, pp. 249-272.

--- (1998) ‘The Functions of Music in Multimedia: A Cognitive Approach’. *In:* S. Yi ed. *Music, Mind, and Science*. Seoul: Seoul National University Press, pp. 53-69.

--- (1993) ‘Associationism and musical soundtrack phenomena’. *Contemporary Music Review,* 9 (1-2), pp. 163-178.

Coulter, J. (2010) ‘Electroacoustic Music with Moving Images: the art of media pairing’. *Organised Sound,* 15 (01), pp. 26-34.

Courribet, B. (2007) ‘Composing Audio-Visual Art: The Issue of Time and the Concept of Transduction’. *Sound and Music Computing Conference.* Marseille: SMC.

Dannenberg, R.B. (2005) ‘Interactive Visual Music: A Personal Perspective’. *Computer Music Journal,* 29 (4), pp. 25-35.

Davis, J.W. (1979) ‘A Response to W. Garner's Observations on the Relationship between Colour and Music’. *Leonardo,* 12 (3), pp. 218-219.

DeWitt, T. (1987) ‘Visual Music: Searching for an Aesthetic’. *Leonardo,* 20 (2, Special Issue: Visual Art, Sound, Music and Technology), pp. 115-122.

Doornbusch, P. (2003) ‘Composers' views on mapping in algorithmic composition’. *Organised Sound,* 7 (02), pp. 145-156.

Edmonds, E. & Pauletto, S. (2004) ‘Audiovisual Discourse in Digital Art’. *ACM SIGGRAPH 2004.* Los Angeles: ACM, pp. 116.

Evans, B. (2005) ‘Foundations of a Visual Music’. *Computer Music Journal,* 29 (4), pp. 11-24.

Fahlenbrach, K. (2005) ‘The Emotional Design of Music Videos. Approaches to Audiovisual Metaphors’. *The Journal of Moving Image Studies,* 4 (1), pp. 1-18.

Faulkner, M. (2006) *VJ: Audio-Visual Art and VJ Culture.* London: Laurence King Publishing.

Fels, S., Gadd, A. & Mulder, A. (2002) ‘Mapping Transparency through Metaphor: Towards More Expressive Musical Instruments’. *Organised Sound,* 7 (2), pp. 109-126.

Garner, W. (1978) 'The Relationship between Colour and Music'. *Leonardo,* 11 (3), pp. 225-226.

Garro, D. (2005) 'A glow on Pythagoras' curtain. A composer's perspective on Electroacoustic music with video'. *Proceedings of EMS05 - Sound in Multimedia Contexts.* Montréal: EMS.

Gomes, G. (2009) *JOPS - Java Particle System Library* [Online] Google. Available from: http://code.google.com/p/jops/ [Accessed 11/18, 2011].

Greenwold, S. (2004) *PSystem: Particle Systems Plugin for Processing* [Online] Rensselaer Polytechnic Institute. Available from: http://www.cs.rpi.edu/~cutler/gaudi/gaudi_from_simon/particle/index.html [Accessed 11/16, 2011].

Hess, R. (2007) *The Essential Blender: Guide to 3D Creation with the Open Source Suite Blender.* Amsterdam: Blender Foundation.

Hunt, A. & Kirk, R. (2000) 'Mapping Strategies for Musical Performance'. *In:* M. Wanderley & M. Battier eds. *Trends in Gestural Control of Music*. Paris: IRCAM - Centre Pompidou, pp. 231-258.

Hunt, A., Wanderley, M. & Kirk, R. (2000) 'Towards a Model for Instrumental Mapping in Expert Musical Interaction'. *Proceedings of the 2000 International Computer Music Conference.* San Francisco: International Computer Music Association, pp. 209.

Hunt, A., Wanderley, M.M. & Paradis, M. (2002) 'The importance of parameter mapping in electronic instrument design'. *Proceedings of the 2002 conference on new interfaces for musical expression.* Singapore: National University of Singapore.

Iwamiya, S. & Ozaki, H. (2004) 'Formal Congruency between Image Patterns and Pitch Patterns'. *Proceedings of the 8th International Conference on Music Perception and Cognition.* Evanston: pp. 145.

Jones, R. & Nevile, B. (2005) 'Creating Visual Music in Jitter: Approaches and Techniques'. *Computer Music Journal,* 29 (4), pp. 55-70.

Junker, G. (2006) *Pro OGRE 3D programming.* New York: Apress.

Kane, B. (2007) 'L'Objet Sonore Maintenant: Pierre Schaeffer, sound objects and the phenomenological reduction'. *Organised Sound,* 12 (01), pp. 15-24.

Kennedy, J. & Eberhart, R.C. (1995) 'Particle swarm optimization'. *Proceedings of IEEE international conference on neural networks.* Perth: IEEE pp. 1942.

Kipfer, P., Segal, M. & Westermann, R. (2004) 'UberFlow: a GPU-based particle engine'. *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware.* Los Angeles: ACM, pp. 115.

Kolb, A., Latta, L. & Rezk-Salama, C. (2004) 'Hardware-based simulation and collision detection for large particle systems'. *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware.* Los Angeles: ACM, pp. 123.

Krajcevski, P. & Reppy, J. (2011)"A Declarative API for Particle Systems". *Practical Aspects of Declarative Languages.* Austin: Springer, pp. 130.

Lander, J. (1998) ‘The ocean spray in your face’. *Game Developer,* 5 (7), pp. 11-16.

Latta, L. (2004) ‘Building a million particle system’. *Game Developers Conference.* San Jose: Citeseer.

Lewkowicz, D.J. (1996) ‘Perception of auditory-visual temporal synchrony in human infants’. *Journal of Experimental Psychology,* 22 (5), pp. 1094-1106.

Lipscomb, S.D. (1994) ‘Perceptual judgement of the relationship between musical and visual components in film’. *Psychomusicology,* 13 (1-2), pp. 60-98.

Lipscomb, S.D. & Tolchinsky, D.E. (2005) ‘The Role of Music Communication in Cinema’/ *In:* D. Miell, R. Macdonald & D.J. Hargreaves eds. *Musical Communication.* New York: Oxford University Press, pp. 383-404.

Marshall, S. & Cohen, A. (1988) ‘Effects of Musical Soundtracks on Attitudes toward Animated Geometric Figures’. *Music Perception,* 6 (1), pp. 95-112.

McAllister, D. (2008) *Particle System API* [Online] McAllister, David. Available from: http://www.particlesystems.org/ [Accessed 11/18, 2011].

--- (2000) ‘The design of an API for particle systems’. *University of North Carolina Technical Report TR 00-007*, 1, pp. 1-8.

Meyer, L.B. (1961) *Emotion and Meaning in Music.* 2nd ed. Chicago: University of Chicago Press.

Molholm, S., Ritter, W., Javitt, D.C. & Foxe, J.J. (2004) ‘Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study’. *Cerebral Cortex,* 14 (4), pp. 452-465.

Momeni, A. & Henry, C. (2006) ‘Dynamic Independent Mapping Layers for Concurrent Control of Audio and Video Synthesis’. *Computer Music Journal,* 30 (1), pp. 49-66.

Momeni, A. & Wessel, D. (2003) ‘Characterizing and controlling musical material intuitively with geometric models’ *Proceedings of the 2003 conference on New interfaces for musical expression.* Singapore: National University of Singapore, pp. 54.

Mulder, A., Fels, S.S. & Mase, K. (1997) ‘Mapping virtual object manipulation to sound variation’. *IPSJ Sig Notes,* 97 (122), pp. 63-68.

Oliver, J. & Pickles, S. (2007) ‘Fijuu2: A Game-Based Audio-Visual Performance and Composition Engine’. *Proceedings of the 7th international conference on new interfaces for musical expression.* New York: ACM Press, pp. 430.

Palamar, T. & Keller, E. (2011) *Mastering Autodesk Maya 2012.* Indianapolis: Wiley Publishing Inc.

Peacock, K. (1988) ‘Instruments to Perform Color-Music: Two Centuries of Technological Experimentation’. *Leonardo,* 21 (4), pp. 397-406.

Piché, J. (2003) ‘De la Musique et des Images’. *Circuit: Musiques Contemporaines,* 13 (3), pp. 41-50.

Reas, C. & Fry, B. (2006) ‘Processing: programming for the media arts’. *AI & Society,* 20 (4), pp. 526-538.

Reeves, W. (1983) ‘Particle systems—a technique for modeling a class of fuzzy objects’. *ACM Transactions on Graphics (TOG),* 2 (2), pp. 91-108.

Reynolds, C.W. (1987) ‘Flocks, herds and schools: A distributed behavioral model’. *Proceedings of the 14th annual conference on Computer graphics and interactive techniques.* New York: ACM New York, pp. 25.

Roads, C. (2005) ‘The art of articulation: The electroacoustic music of Horacio Vaggione’. *Contemporary Music Review,* 24 (4), pp. 295-309.

--- (2004) *Microsound.* Cambridge: The MIT Press.

--- (1988) ‘Introduction to granular synthesis’. *Computer Music Journal,* 12 (2), pp. 11-13.

Rovan, J.B., Wanderley, M.M., Dubnov, S. & Depalle, P. (1997) ‘Instrumental gestural mapping strategies as expressivity determinants in computer music performance’. *Kansei, The Technology of Emotion. Proceedings of the AIMI International Workshop.* Citeseer, pp. 68.

Rudi, J. (2005) ‘Computer Music Video: A Composer's Perspective’. *Computer Music Journal,* 29 (4), pp. 36-44.

--- (1998) ‘Computer music animations’. *Organised Sound,* 3 (03), pp. 193-198.

Serafin, S. (2008) ‘Computer Generation and Manipulation of Sounds’. *In:* N. Collins & J. d'Escriván eds. *The Cambridge Companion to Electronic Music*. Cambridge: Cambridge University Press, pp. 203-217.

Sims, K. (1990) *Particle animation and rendering using data parallel computation.* New York: ACM.

Snyder, B. (2000) *Music and Memory: An Introduction.* Cambridge: MIT Press.

Sommerer, C. & Mignonneau, L. (1999) ‘Art as a Living System: Interactive Computer Artworks’. *Leonardo,* 32 (3), pp. 165-173.

Spence, C. (2007) ‘Audiovisual multisensory integration’. *Acoustical Science and Technology,* 28 (2), pp. 61-70.

Spence, C. & Squire, S. (2003) ‘Multisensory integration: maintaining the perception of synchrony’. *Current Biology,* 13 (13), pp. 519-521.

Steenhuisen, P. (2009) *Sonic mosaics: conversations with composers.* Edmonton: University of Alberta.

Steiner, H.C. (2008) *The [hid] Toolkit for Pd.* Masters thesis. New York: New York University.

--- (2005) ‘[hid] toolkit: a unified framework for instrument design’. *Proceedings of the 2005 conference on new interfaces for musical expression.* Singapore: National University of Singapore, pp. 140.

Steiner, H.C. & Henry, C. (2007) ‘Progress report on the mapping library for pd’. *Proceedings of the PureData Convention.* Montreal: IEM.

Stockhausen, K. (1959) ‘... How Time Passes ...’. *Die Reihe,* 3, pp. 10-43.

Strick, J. (2005) ‘Visual Music’. *In:* K. Brougher & J. Zilzcer eds. *Visual Music: Synaesthesia in Art and Music Since 1900* New York: Thames and Hudson, pp. 15-20.

Sugita, Y. & Suzuki, Y. (2003) ‘Audiovisual perception: Implicit estimation of sound-arrival time’. *Nature,* 421 (6926), pp. 911-912.

Toiviainen, P. (2008) ‘The Psychology of Electronic Music’. *In:* N. Collins & J. d'Escriván eds. *The Cambridge Companion to Electronic Music*. Cambridge: Cambridge University Press, pp. 218-231.

Tolonen, T., Välimäki, V. & Karjalainen, M. (1998) *Evaluation of modern sound synthesis methods.* Helsinki: Helsinki University of Technology.

Trenholme, D. & Smith, S.P. (2008) ‘Computer game engines for developing first-person virtual environments’. *Virtual Reality,* 12 (3), pp. 181-187.

Truax, B. (1988) ‘Real-time granular synthesis with a digital signal processor’. *Computer Music Journal,* 12 (1), pp. 14-26.

Truax, B. (1990) ‘Composing with Real-Time Granular Sound’. *Perspectives of New Music,* 28 (2), pp. 120-134.

Vaggione, H. (2001) ‘Some ontological remarks about music composition processes’. *Computer Music Journal,* 25 (1), pp. 54-61.

Van Nort, D. & Wanderley, M.M. (2006) ‘The LoM Mapping Toolbox for Max/MSP/Jitter’. *Proceedings of the International Computer Music Conference.* New Orleans: ICM, .

Welch, R.B. & Warren, D.H. (1980) ‘Immediate perceptual response to intersensory discrepancy’. *Psychological Bulletin,* 88 (3), pp. 638-667.

Wells, A. (1980) ‘Music and Visual Color: A Proposed Correlation’. *Leonardo,* 13 (2), pp. 101-107.

Whitney, J. (1994) 'To Paint on Water: The Audiovisual Duet of Complementarity'. *Computer Music Journal,* 18 (3), pp. 45-52.

--- (1991) ‘Fifty Years of Composing Computer Music and Graphics: How Time's New Solid-State Tractability Has Changed Audio-Visual Perspectives’. *Leonardo,* 24 (5), pp. 597-599.

--- (1980) *Digital Harmony: On the Complementarity of Music and Visual Art.* Michigan: Byte Books.

Wishart, T. (1996) *On Sonic Art.* Amsterdam: Harwood Academic Publishers.

--- (1986) ‘Sound Symbols and Landscapes’. *In:* S. Emmerson ed. *The Language of Electroacoustic Music*. London: Macmillan, pp. 41-60.

Wolek, N. (2002) ‘A Granular Toolkit for Cycling74's Max/MSP’. *SEAMUS 2002 National Conference.* Iowa City: SEAMUS.

Wright, M. (2005) ‘Open sound control: an enabling technology for musical networking’. *Organised Sound,* 10 (03), pp. 193-200.

Wright, M. & Freed, A. (1997) ‘Open sound control: A new protocol for communicating with sound synthesizers’. *Proceedings of the 1997 International Computer Music Conference.* Thessaloniki: ICM, pp. 101.

Wright, M., Freed, A. & Momeni, A. (2003) ‘Open Sound Control: State of the art 2003’. *Proceedings of the 2003 conference on new interfaces for musical expression.* Singapore: National University of Singapore, pp. 153.

Xenakis, I. & Brown, R. (1989) ‘Concerning time’. *Perspectives of New Music,* 27 (1), pp. 84-92.

Xith3D (2010) *Xith3D Project* [Online] Xith3D. Available from: http://xith.org/ [Accessed 18/11, 2011].

Zilczer, J. (2005) 'Music for the Eyes: Abstract Painting and Light Art'. *In:* K. Brougher & J. Zilczer eds. *Visual Music: Synaesthesia in Art and Music Since 1900* New York: Thames and Hudson, pp. 25-85.

--- (1987) ‘Color Music: Synaesthesia and Nineteenth-Century Sources for Abstract Art’. *Artibus et Historiae,* 16 (1), pp. 101-126.

Appendix I

Practical Submission

The data DVD included with this thesis comprises the practical component of the research. The disc contains all documented software systems and video works, with source and generated audio material included where relevant. All video media is in high definition 720p format and uses the AVC and AAC codecs for video and audio respectively. Source and generated audio material is in 24 bit uncompressed Wave format with a sample rate of 44.1kHz. The included *Java* libraries and their implementations have been tested with *Processing* version 1.5.1, while the granular synthesis systems have been tested with *Max/MSP* version 5.1.7.

The following table outlines the folder structure of the disc:

07 – Composition Studies	
7.2 – Io	
Audio	Source and generated audio material used throughout the creation of *Io*.
Io_720p_AVC.mp4	The final video file for *Io*.
7.3 – Rhea	
Audio	Source and generated audio material used throughout the creation of *Rhea*.
Rhea_720p_AVC.mp4	The final video file for *Rhea*.
7.4 – Dione	
Audio	Generated audio material used throughout the creation of *Dione*.
Dione_720p_AVC.mp4	The final video file for *Dione*.
08 – Mapping Studies	
8.2 – Mapping Study 1	
Audio	Source and generated audio material used throughout the creation of *Mapping Study 1.*
MappingStudy1_720p_AVC.mp4	The final video file for *Mapping Study 1*.
8. 3 – Mapping Study 2	
Audio	Source and generated audio material used throughout the creation of *Mapping Study 2.*
MappingStudy2_720p_AVC.mp4	The final video file for Mapping Study 2.
8. 4 – Mapping Study 3	
Audio	Source and generated audio material used throughout the creation of *Mapping Study 3.*
MappingStudy3_720p_AVC.mp4	The final video file for *Mapping Study 3*.
8.5 – Mapping Study 4	
Audio	Source and generated audio material used throughout the creation of *Mapping Study 4.*
MappingStudy4a_720p_AVC.mp4	The final video file for the first part of *Mapping Study 4*.
MappingStudy4b_720p_AVC.mp4	The final video file for the second part of *Mapping Study 4*.
09 – In Perpetuity	
9.2 – Software Tools	
Pmlib	The *pmlib* Java library including source code, documentation and example usage.
psl.granular	The *psl.granular* package including the *Max/MSP* files for each component system.
Pslib	The *pslib* Java library including source code and documentation.
9.4 – The Early Lilacs	
Audio	Source and generated audio material used throughout the creation of *The Early Lilacs.*
Data	Source and generated data used throughout the creation of *The Early Lilacs*.
Software	The *Processing* sketches used throughout the creation of *The Early lilacs*.
TheEarlyLilacs_720p_AVC.mp4	The final video file for *In Perpetuity: The Early Lilacs*.
9.5 – The Linden Trees	
Audio	Source and generated audio material used throughout the creation of *The Linden Trees.*
Data	Source and generated data used throughout the creation of *The Linden Trees*.
Software	The *Processing* sketches used throughout the creation of *The Linden Trees*.
TheLindenTrees_720p_AVC.mp4	The final video file for *In Perpetuity: The Linden Trees*.

Appendix II

In Perpetuity: The Early Lilacs – Media Generation Data

Source Audio File	Preset	Source Data	Pit. Scl.	Dur. Scl.	Output Audio File	Length	Format	Sync. Time
*Pre-Granualisation**								
city_01_44k24m.wav	city1_g.txt	n/a	4.00	n/a	city1_g1_44k24m.wav	115s	44k24m	n/a
city_01_44k24m.wav	city1_g.txt	n/a	8.00	n/a	city1_g2_44k24m.wav	154s	44k24m	n/a
city_01_44k24m.wav	city1_g.txt	n/a	4.00	n/a	city1_g3_44k24m.wav	66s	44k24m	n/a
city_02_44k24m.wav	city2_g.txt	n/a	16.00	n/a	city2_g1_44k24m.wav	81s	44k24m	n/a
city_02_44k24m.wav	city2_g.txt	n/a	6.00	n/a	city2_g2_44k24m.wav	155s	44k24m	n/a
city_02_44k24m.wav	city2_g.txt	n/a	16.00	n/a	city2_g3_44k24m.wav	94s	44k24m	n/a
school_1b_44k24m.wav	sch1b_g.txt	n/a	2.00	n/a	sch1b_g1_44k24m.wav	323s	44k24m	n/a
school_1d_44k24m.wav	sch1d_g.txt	n/a	2.00	n/a	sch1d_g1_44k24m.wav	213s	44k24m	n/a
school_1d_44k24m.wav	sch1d_g.txt	n/a	8.00	n/a	sch1d_g2_44k24m.wav	165s	44k24m	n/a
school_1d_44k24m.wav	sch1d_g.txt	n/a	0.25	n/a	sch1d_g3_44k24m.wav	158s	44k24m	n/a
school_1e_44k24m.wav	sch1e_g.txt	n/a	16.00	n/a	sch1e_g1_44k24m.wav	62s	44k24m	n/a
Grain List								
city1_g1_44k24m.wav	n/a	fs01a_g.seq	1.50	1.00	fs01a_g1_44k24s.wav	139s	44k24s	00:05.200
city2_g1_44k24m.wav	n/a	fs01a_g.seq	8.00	1.00	fs01a_g2_44k24s.wav	139s	44k24s	00:05.200
city1_g2_44k24m.wav	n/a	fs01a_g.seq	4.00	1.00	fs01a_g3_44k24s.wav	140s	44k24s	00:05.200
city1_g3_44k24m.wav	n/a	fs01c_g.seq	2.00	1.00	fs01c_g1_44k24s.wav	109s	44k24s	04:12.020
city2_g2_44k24m.wav	n/a	fs01c_g.seq	4.00	1.00	fs01c_g2_44k24s.wav	109s	44k24s	04:12.020
city2_g3_44k24m.wav	n/a	fs01c_g.seq	16.00	10.00	fs01c_g3_44k24s.wav	111s	44k24s	04:12.020
city_01_44k24m.wav	n/a	fs01fx1_g.seq	1.00	1.00	fs01fx1_g1_44k24m.wav	50s	44k24m	03:29.580
city_01_44k24m.wav	n/a	fs01fx1_g.seq	8.00	1.00	fs01fx1_g2_44k24m.wav	48s	44k24m	03:29.580
Emitter Analysis								
school_1d_44k24m.wav	fs01a_e.txt	fs01a_e.seq	1.50	n/a	fs01a_e1_44k24s.wav	162s	44k24s	00:05.420
school_1c_44k24m.wav	fs01a_e.txt	fs01a_e.seq	2.00	n/a	fs01a_e2_44k24s.wav	163s	44k24s	00:05.420
city_02_44k24m.wav	fs01a_e.txt	fs01a_e.seq	6.00	n/a	fs01a_e3_44k24s.wav	164s	44k24s	00:05.420
city_02_44k24m.wav	fs01a_e.txt	fs01a_e.seq	0.50	n/a	fs01a_e4_44k24s.wav	163s	44k24s	00:05.420
city_02_44k24m.wav	fs01c_e.txt	fs01c_e.seq	4.00	n/a	fs01b_e1_44k24s.wav	145s	44k24s	00:05.420
city_02_44k24m.wav	fs01c_e.txt	fs01c_e.seq	6.00	n/a	fs01b_e2_44k24s.wav	147s	44k24s	00:05.420
water_01_44k24m.wav	fs01c_e.txt	fs01c_e.seq	16.00	n/a	fs01b_e3_44k24s.wav	154s	44k24s	00:05.420
school_1e_44k24m.wav	fs01c_e.txt	fs01c_e.seq	4.00	n/a	fs01b_e4_44k24s.wav	150s	44k24s	00:05.420
Data Driven								
city_01_44k24m.wav	fs01b_d.txt	fs01b.seq	8.00	n/a	fs01b_d1_44k24s.wav	141s	44k24s	02:12.900
city_01_44k24m.wav	fs01b_d.txt	fs01b.seq	2.00	n/a	fs01b_d2_44k24s.wav	161s	44k24s	02:12.900
city_01_44k24m.wav	fs01b_d.txt	fs01b.seq	0.75	n/a	fs01b_d3_44k24s.wav	183s	44k24s	02:12.900
school_1a_44k24m.wav	fs01b_d.txt	fs01b.seq	4.00	n/a	fs01b_d4_44k24s.wav	175s	44k24s	02:12.900
school_1a_44k24m.wav	fs01b_d.txt	fs01b.seq	0.50	n/a	fs01b_d5_44k24s.wav	168s	44k24s	02:12.900
school_1b_44k24m.wav	fs01c_d.txt	fs01c.seq	1.00	n/a	fs01c_d1_44k24s.wav	148s	44k24s	04:12.020
school_1b_44k24m.wav	fs01c_d.txt	fs01c.seq	4.00	n/a	fs01c_d2_44k24s.wav	144s	44k24s	04:12.020
school_1b_44k24m.wav	fs01c_d.txt	fs01c.seq	0.25	n/a	fs01c_d3_44k24s.wav	145s	44k24s	04:12.020

** All pre-granualisation audio was also rendered in 44k24s format to provide unmapped material*

Appendix III

In Perpetuity: The Linden Trees – Media Generation Data

Source Audio File	**Preset**	**Source Data**	**Pit. Scl.**	**Dur. Scl.**	**Output Audio File**	**Length**	**Format**	**Sync. Time**
*Pre-Granualisation**								
park_01e1_44k24m.wav	park_g1.txt	n/a	8.00	n/a	park_g1_44k24m.wav	148s	44k24m	n/a
park_01e1_44k24m.wav	park_g2.txt	n/a	8.00	n/a	park_g2_44k24m.wav	97s	44k24m	n/a
park_01e2_44k24m.wav	park_g3.txt	n/a	2.00	n/a	park_g3_44k24m.wav	128s	44k24m	n/a
park_01e2_44k24m.wav	park_g4.txt	n/a	4.00	n/a	park_g4_44k24m.wav	78s	44k24m	n/a
park_02e1_44kw4m.wav	park_g5.txt	n/a	4.00	n/a	park_g5_44k24m.wav	115s	44k24m	n/a
park_02e1_44kw4m.wav	park_g6.txt	n/a	8.00	n/a	park_g6_44k24m.wav	93s	44k24m	n/a
park_02e2_44k24m.wav	park_g7.txt	n/a	8.00	n/a	park_g7_44k24m.wav	139s	44k24m	n/a
park_02e2_44k24m.wav	park_g8.txt	n/a	4.00	n/a	park_g8_44k24m.wav	125s	44k24m	n/a
school_1a_44k24m.wav	school_g1.txt	n/a	3.00	n/a	school_g1_44k24m.wav	126s	44k24m	n/a
school_1a_44k24m.wav	school_g2.txt	n/a	6.00	n/a	school_g2_44k24m.wav	113s	44k24m	n/a
Grain List								
park_g1_44k24m.wav	n/a	fs02a_g.seq	1.00	1.50	fs02a_g1_44k24s.wav	160s	44k24s	00:09.280
park_g4_44k24m.wav	n/a	fs02a_g.seq	1.00	1.50	fs02a_g2_44k24s.wav	158s	44k24s	00:09.280
school_g1_44k24m.wav	n/a	fs02a_g.seq	0.50	1.50	fs02a_g3_44k24s.wav	159s	44k24s	00:09.280
park_g2_44k24m.wav	n/a	fs02b_g.seq	1.00	1.00	fs02b_g1_44k24s.wav	99s	44k24s	03:40.300
park_g6_44k24m.wav	n/a	fs02b_g.seq	0.50	1.00	fs02b_g2_44k24s.wav	90s	44k24s	03:40.300
school_g2_44k24m.wav	n/a	fs02b_g.seq	0.25	1.00	fs02b_g2_44k24s.wav	92s	44k24s	03:40.300
Emitter Analysis								
park_01e1_44k24m.wav	fs02a_e1.txt	fs02a_e.seq	6.00	n/a	fs02a_e1_44k24s.wav	160s	44k24s	00:09.280
park_03e1_44k24m.wav	fs02a_e2.txt	fs02a_e.seq	3.00	n/a	fs02a_e2_44k24s.wav	165s	44k24s	00:09.280
schoo1_1b_44k24m.wav	fs02a_e3.txt	fs02a_e.seq	2.00	n/a	fs02a_e3_44k24s.wav	159s	44k24s	00:09.280
park_03e2_44k24m.wav	fs02b_e1.txt	fs02b_e.seq	2.00	n/a	fs02b_e1_44k24s.wav	85s	44k24s	02:20:540
park_04e1_44k24m.wav	fs02b_e2.txt	fs02b_e.seq	8.00	n/a	fs02b_e2_44k24s.wav	125s	44k24s	02:20:540
school_1c_44k24m.wav	fs02b_e3.txt	fs02b_e.seq	4.00	n/a	fs02b_e3_44k24s.wav	104s	44k24s	02:20:540
park_01e1_44k24m.wav	fs02c_e1.txt	fs02c_e.seq	2.00	n/a	fs02c_e1_44k24s.wav	164s	44k24s	03:40.300
school_1e_44k24m.wav	fs02c_e2.txt	fs02c_e.seq	3.00	n/a	fs02c_e2_44k24s.wav	161s	44k24s	03:40.300
school_1d_44k24m.wav	fs02c_e3.txt	fs02c_e.seq	1.50	n/a	fs02c_e3_44k24s.wav	162s	44k24s	03:40.300
Data Driven								
park_04e1_44k24m.wav	fs02a_d1.txt	fs02a_cv.seq	3.00	n/a	fs02a_d1_44k24s.wav	168s	44k24s	00:09.280
park_03e1_44k24m.wav	fs02a_d2.txt	fs02a_cv.seq	4.00	n/a	fs02a_d2_44k24s.wav	157s	44k24s	00:09.280
school_1a_44k24m.wav	fs02a_d3.txt	fs02a_cv.seq	2.00	n/a	fs02a_d3_44k24s.wav	163s	44k24s	00:09.280
park_03e1_44k24m.wav	fs02b_d1.txt	fs02b_cv.seq	2.00	n/a	fs02b_d1_44k24s.wav	90s	44k24s	02:20:540
park_01e2_44k24m.wav	fs02b_d2.txt	fs02b_cv.seq	3.00	n/a	fs02b_d2_44k24s.wav	84s	44k24s	02:20:540
school_1b_44k24m.wav	fs02b_d3.txt	fs02b_cv.seq	0.75	n/a	fs02b_d3_44k24s.wav	98s	44k24s	02:20:540
park_01e1_44k24m.wav	fs02c_d1.txt	fs02c_cv.seq	1.50	n/a	fs02c_d1_44k24s.wav	158s	44k24s	03:40.300
park_03e2_44k24m.wav	fs02c_d2.txt	fs02c_cv.seq	4.00	n/a	fs02c_d2_44k24s.wav	160s	44k24s	03:40.300
school_1e_44k24m.wav	fs02c_d3.txt	fs02c_cv.seq	3.00	n/a	fs02c_d3_44k24s.wav	166s	44k24s	03:40.300

** All pre-granualisation audio was also rendered in 44k24s format to provide unmapped material*