Audiovisual Particles

Mapping strategies for a perceptual correlation between visual particles and audio grains

1. INTRODUCTION

Parameter mapping strategies arguably provide a conceptual framework for the creation of audiovisual art in which a correspondence between audio and video mediums is sought. This paper provides an overview of parameter mapping discourse and identifies prevalent terminology relevant to the creation of audiovisual mapping strategies. Organisational time scale is discussed as a basis for deriving parallels between structures within both aural and visual mediums, allowing the conceptual objects *particle* and *cloud* to be defined. Following the identification of relevant terminology, the works *Cirrus* and *Photuris* are presented as example mapping strategies and the relevance of parameter mapping within audiovisual composition is discussed.

2. PARAMETER MAPPING

It is arguable that the process of parameter mapping – a transposition of values between domains – may offer a compositional framework for the creation of audiovisual art. While parameter mapping is commonly discussed within the context of musical instrument design, mapping theory within the field of audiovisual composition is rare. In discussing algorithmic composition, Doornbusch (2002:145) notes due to the fundamental integration of mapping systems within the compositional process, parameter mapping is rarely explicitly discussed. The same may be observed within the field of audiovisual composition; the process of mapping values between mediums is rarely seen as a discrete stage of practice within the creation of audiovisual works. Literature within the field of instrument design, and algorithmic composition may, however, offer terminology and techniques relevant to the creation of audiovisual works in which a correspondence between mediums is sought.

2.1. Audiovisual Correspondence

Since the early twentieth century artists have sought correlations between the properties of both aural and visual mediums. (Alexander and Collins, 2008) An area of sustained focus is that of a perceptual correspondence between colour and pitch, based on either a physical correlations between the waveforms of each medium, or derived through perceived synaesthetic correlations. (Alves, 2005; Davis, 1979; DeWitt, 1987; Garner, 1978; Wells, 1980; Zilczer, 1987; Zilcer, 2005) Other notable mappings between mediums include links between visual form and aural timbre, visual density and aural polyphony, and visual position to aural pitch and amplitude. (Bailey *et al.* 2007; Düchting, 2004; Galeyev, 1976; Jones and Neville, 2005) Regardless of the validity of such perceptual correlations, it is clear that for cross-modal correspondences to be perceived, the compositional process must adopt a system of mappings between attributes within each medium. Such mappings may be derived through artistic interpretation, computer analysis, generative or algorithmic systems or any combination of approaches. (Abbado, 1988; Edmonds *et al.*, 2004; Edmonds and Pauletto, 2004; Gerhard, 1999) It may therefore be argued that audiovisual composition that seeks a perceptual correspondence between sound and image is based on the creation and subversion of parameter mappings whether explicit or implied. (Alves, 2005; Dannenberg, 2005; Jones and Neville, 2005; Rudi, 2005)

2.2. Mapping Types

It is argued that two types of mapping strategy exist; explicit and implicit. Arfib *et al.* (2002) define explicit mappings as those in which 'one can exactly describe the links between the input and the output mapping parameters.' in contrast to implicit mappings in which 'the mapping box is considered as a black box for which we define behaviour rules but not precise values.' A distinction is also made between dynamic and static mapping behaviours. A dynamic mapping behaviour is defined as a mapping with the ability to evolve in time. Conversely a static mapping behaviour will not adapt to or learn from input data. Momeni and Henry (2006:50) define a dynamic mapping layer as one in which 'the internal behaviour of the system can produce output variation without variation in the input' through time-variable behaviour. Mapping strategies for instrument design generally fall into one of three categories: One-to-one, one-to-many, and many-to-one. (Fels *et al.*, 2002) Through combination, many-to-many mappings may be formed and these will often prove more satisfactory for both performer and audience after a period of learning. Indeed, Hunt and Kirk (2002:255) note that the multi-parametric (many-to-many) interface tested within their research consistently allowed the users to improve over time and, as such, proved to be the most engaging interface within their tests.

Following interviews with a selection of algorithmic composers, Doornbusch (2002:155) notes that complex mappings appear to be more common within the works of the composers interviewed. This assertion is based on the observation that simple mapping systems are often less successful unless the data to be mapped is itself complex. Doornbusch notes the similarity between his observations and those made by Hunt and Kirk (2000:255) and offers that, in this instance, theory from the field of instrument design translates into the field of algorithmic composition. A similar parallel may be observed within the field of audiovisual composition: Dannenberg (2005:28) notes that while simple correspondences may be initially appealing, they rapidly cease to be interesting, while Jones and Neville (2005:57) offer that simple linear mappings based on synaesthetic association may be uninteresting in isolation. As such, complex or dynamic mappings within audiovisual composition may ultimately prove more satisfactory than simple one-to-one mappings.

2.3. Mapping Transparency

Fels *et al.* offer the term 'transparency' as terminology to provide 'an indication of psychophysiological distance, in the minds of the player and the audience, between the input and output of a device mapping.' (2002:109) On this basis, a fully transparent mapping is one in which both audience and performer derive a connection between parameter in different mediums, without the need for explanation or instruction. It is argued that metaphor and culturally-derived associations facilitate the creation of transparent mappings. (Fels et al. 2002:109) Indeed Alves (2005:47) notes 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.' Similarly, Jones and Neville (2005:56) observe that a mapping between aural amplitude and visual brightness may appear natural due to both parameters being a measure of physical intensity within their respective mediums. As such, it may be argued that audiovisual mappings that have a basis within the physical world map appear more transparent.

2.4. Mapping Levels

A mapping level may be defined as an interpretative function that exists between the input and output parameters of a mapping system. (Momeni and Henry, 2006:50) Through the accumulation of mapping levels, strategies of increasing complexity may be formed. It is arguable that two types of multi-tiered mapping strategy exist; serial and parallel. (Figure 2.1) In a serial mapping strategy each layer outputs directly to the next interpretive level and is subject to the interpretation that preceded it. On this basis, the connection between input and output parameters of the system is derived through the accumulation of parameter relationships. The simplest form of parallel mapping strategy is equivalent to a one-to-many mapping system; a single set of input parameters is split amongst multiple interpretive layers. On this basis, each output parameter is subject only to the mapping layer between itself and the input. Through the combination of serial and parallel mapping systems it is possible to create strategies of varying complexity. Within such as system it should be possible to trace the accumulated effects of each mapping level although it is unlikely that this will be apparent to an audience.

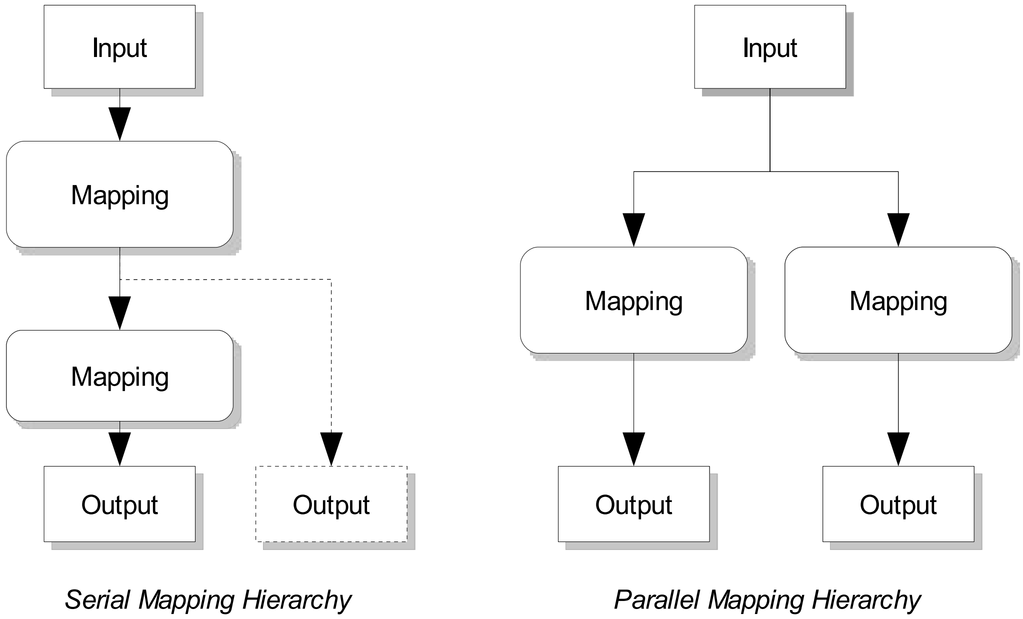

Figure 2.1. Block diagrams showing the flow of data from input to output in serial and parallel mapping hierarchies.

3. AUDIOVISUAL PARTICLES

3.1. Compositional Time Scales

Compositional structure may be divided into discrete organisational levels that exist within specific time scales. Each structural layer within a composition will contain numerous parameters that may be subject to a temporal morphology. On this basis, any morphology within musical structure may be said to exist within a specific time scale. Of primary relevance to this paper are the micro, sound object, meso and macro time scales.

Roads (2004) defines the micro time scale as a scale dealing with sound events ranging from several hundred microseconds to around one hundred milliseconds in duration; termed by Roads as 'transient audio phenomena.' The organisation of such microevents into discrete structures allows the formation of sonic 'clouds.' Such structures may be formed through the process of granular synthesis – the segmentation of audio material into micro-level grains and subsequent re-assembly by a synthesis algorithm. (Truax, 1988, 1990; Roads, 2004:96) Roads notes that 'clouds encourage a process of statistical evolution. Within this evolution the composer can impose specific morphologies.' (Roads, 2004:15) Dependent on a duration between 100ms and a few seconds, cloud structures may be termed heterogeneous sound objects. (Blackwell and Young, 2004:126; Roads 2004:19) Such objects discard the static homogeneous characteristics of traditional musical note, instead incorporating time-variant properties, a feature described by Wishart (1996:93) as the 'dynamic morphology' of the object. Cloud structures with a morphology of greater duration, typically measured in seconds, exist within the meso time-scale. Such structures may be textural in nature (Roads, 2004:15) or may be the combinatorial product of numerous sound object-level structures. (Blackwell and Young, 2004:126) The subsequent organisation of meso-level structures into musical form make up the macrostructure of the composition.

3.2. Audiovisual Particles

Within the field of computer graphics, particle systems may be used to render fuzzy objects such as smoke or fire. (Reeves, 1983) Such a system will be subject to organisation at a range of structural levels. As such, the underlying structure of a particle system is analogous to the structural composition of audio clouds. Within the micro time scale exists the visual particle; a infinitesimally small object within a virtual space, to which colour and/or image may be assigned. According to the durations defined by Roads (2004), the organisation of such particles into cloud formations may be deemed sound object or meso-level structures. It is clear, therefore, that strong parallels exist between the formation of audio clouds and visual particle systems, and that four compositional times scales exist, within which both audio and visual structures may be organised. Similarly it appears possible for a compositional approach within one medium to be successfully reproduced within the other, subject to an interpretative process. It is arguable, therefore, that the term *particle* may be used as descriptor for a micro-level object that exists within either the visual or aural medium. This theory assumes no pre-determined connections between microstructures in either medium. Such terminology may provide a conceptual basis for the transposition of organisational strategies between mediums and may facilitate the formation of cross-modal mapping strategies within the defined compositional time-scales. Within the context of this research, it is assumed that an audio particle may be rendered as a sound grain, and a video particle may be rendered as a visual object within a virtual space, to which colour and/or image may be assigned. Further to this, the term *cloud* may be applied to any particle structure that is subject to organisation within the sound object or meso time-scales.

4. MAPPING STRATEGIES

Within the context of this research, an audiovisual mapping strategy may be considered a conceptual framework for audiovisual composition. Such a strategies may take many different forms, ranging from linear perceptual correspondences between mediums to complex dynamic mapping systems. A distinction must be drawn between mapping strategies that seek to interpret the parameters of one medium to the other, and those that utilise a single data set to simultaneously control events within each medium. It may be assumed that an audiovisual mapping strategy may inheret terminology from the fields of instrument design, and algorithmic composition. As such, mapping strategies may be implicit or explicit, simple or complex, linear or dynamic. Similarly the perceptual strength of mappings within a strategy may discussed in terms of transparency or consonance and dissonance, with opaque mappings being dissonant and transparent mappings consonant. A mapping strategy may incorporate multiple interpretive levels that exist within the defined compositional time scales.

4.1. Mapping Time Scales

Each layer of an audiovisual mapping strategy exists within one of the established compositional time scales. On this basis, a mapping between visual particle position and grain pitch may be deemed a micro-level mapping for example, whilst a mapping between visual cloud density and aural cloud density may be considered a sound object-level or meso-level strategy, dependent on duration. It would appear possible that a mapping may transcend time scales, an example of this being a mapping between visual cloud brightness and individual grain timbre. While this theory is applicable on a conceptual level, the effects of such mappings may be more apparent within the highest level time scale. In the mapping above for example, it is likely that the accumulated effect of multiple grains being modified by the meso-level parameter of cloud brightness will be perceived as a meso-level to meso-level mapping between cloud brightness and cloud timbre. It is clear that in this case, the mapping between cloud and grain will not be easily perceived, it is therefore arguable, that mappings that reside within a compositional time scale will be more transparent than mappings between different level structures.

4.2. Cirrus

Cirrus (2009) is a linear audiovisual composition that explores perceptual mappings between visual particle systems and granular synthesis. All mappings within Cirrus are purely perceptual – they are based on artistic interpretation between mediums rather than the computer analysis of attributes within mediums. Cirrus incorporates an *image-sourced particle system* for the creation of visuals. This is a visual instrument that derives particle co-ordinate data from digital imagery through a combination of thresholding and degradation. This data is then stored within a set of maps that may be controlled in real-time, thus allowing the formation of complex visual imagery through the accumulation of particle trajectories. Audio for the composition was generated through a process of 'quasi-synchronous granular synthesis,' (Roads, 2004:93) using found sounds as the source material.

The mapping strategy within Cirrus combined simple linear mappings with artistic decisions based on aesthetic appearance. All of the mapped elements within the composition take the form of strict one-to-one and one-to-many mappings using data derived through artistic perception of attributes within the video medium. The primary mapping strategy is that of a transposition of perceived video attributes onto the parameters of multiple sonic textures – grain clouds within the meso time-scale. There are two primary mappings within Cirrus, both of which operate within the meso time-scale: Perceived visual brightness was mapped to aural timbre through a combination of filtering, texture density and audio sample selection, while perceived visual density was mapped to audio texture amplitude. It should be noted that while the mapping strategy was very strict throughout the initial stages of the composition process, latter experimentation showed that accentuation of the core mappings within the macro time-scale appeared to improve the dynamic range of the piece and helped alleviate some of the predictability that occurred through the mapping process. Similarly, to provide interest and to create a sense of audiovisual counterpoint, (Galeyev, 1976; Whitney, 1980) many unmapped elements where incorporated. Observations made throughout the compositional process highlight the importance of a perceptual audiovisual consonance or dissonance and it is apparent that the inclusion of unmapped elements goes some way to maintain interest when the mapping strategy used lacks complexity.

4.3. Photuris

In contrast to the implicit mapping system within Cirrus, Photuris (2009) is an experiment into strict sound object and meso-level mappings derived from computer analysis of visual imagery. Photuris also explores behavioural simulation as a data set to control the temporal morphology of mediums.

The use of a separate data set for the concurrent synthesis of audio and visual data is explored by Momeni and Henry (2006). While their research details solutions for the gestural control of synthesis parameters many of their observations are relevant within the creation of non-interactive works. Momeni and Henry propose that the process of concurrent synthesis derived from a single data set provides an alternative approach to methods that take a lead medium and derive control variables through analytical processes. Similarly, they argue that a concurrent one-to-many mapping system allows for complete cross-modal synchrony which may then be abstracted as required. Research by Blackwell and Young (2004) also explores parameter mapping strategies driven by a single data set. In this work the generation of audio through behavioural simulation is discussed with specific relevance to the swarming algorithms proposed by Reynolds. (1987)

Similarly, Momeni and Wessell (2003) present 'Boids Space,' an implementation of flocking simulation within a live performance instrument.

Photuris uses an modified version of the image-sourced particle system developed for the creation of Cirrus. The updated particle system incorporates Reynolds' flocking algorithm (1987) to allow both target-seeking behaviour based on image analysis, and swarm behaviour based on the stigmergetic interactions within the behavioural simulation. Figure 4.1 details the mapping hierarchy used within the creation of Photuris. Unlike the concurrent audiovisual synthesis processes within Momeni and Henry's work (2006), the mapping strategy employed is serial in nature: Boid (Reynolds, 1987) position and velocity data are mapped to visual particle parameters, the attributes of which are then mapped to granular synthesis parameters following analysis of the visual output. This strategy means that while the granular synthesis properties are ultimately based on flocking data, the values are subject to two levels of interpretation. It is arguable therefore, that mappings between simulation and video are more transparent than those between simulation and audio. While such a discrepancy between mapping transparencies may be apparent within the context of instrument design, it is arguable that such a discrepancy is of less concern within audiovisual composition, where the controlling data set need not be apparent to the audience.

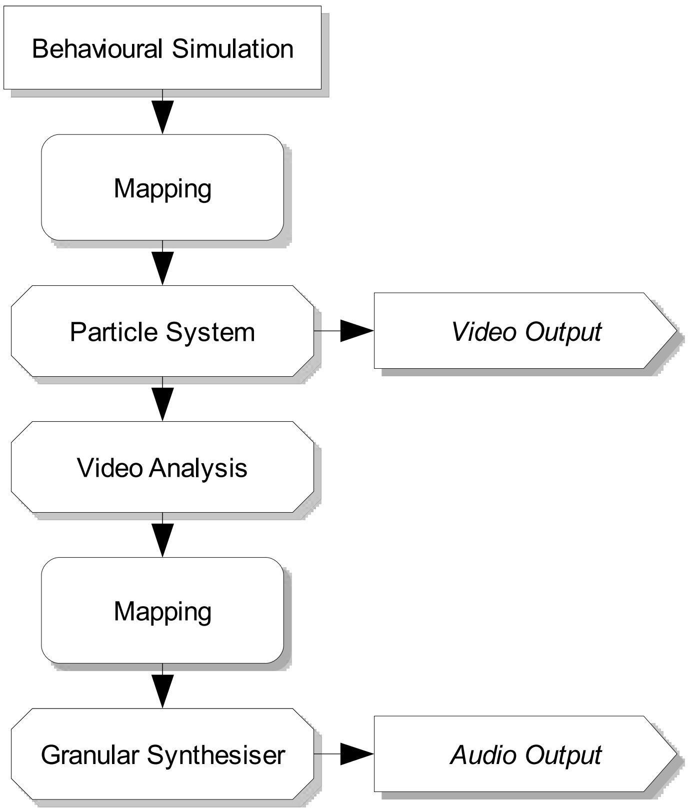

Figure 4.1. Block diagram showing the data flow between behavioural simulation and audio/video synthesisers.

Due to the processing implications within real-time flocking simulation (Reynolds, 1987) it was not possible to calculate the flock data for each particle. Instead, each particle has a parent boid, from which it derives position and velocity data. As such, each boid controls a cloud of visual particles that form a meso-level structure. Correspondences between visual and aural mediums are derived from meso-level mappings between visual cloud data and granular synthesis parameters. At a perceptual level mappings within Photuris are based on macro-level correlations between visual brightness and aural timbre, and visual density and aural amplitude. The meso-level mapping strategy required to achieve this was derived through experimentation: Audio cloud density is varied based on the average brightness of each cloud, while audio cloud amplitude is based on visual cloud spread. Sound object-level mappings between visual cloud separation and aural pitch variation are also incorporated. Similar to the compositional process employed in the creation of Cirrus, unmapped sonic elements are incorporated to create an audiovisual counterpoint and to maintain interest throughout the duration of the piece. A notable observation made throughout the composition of Photuris is that the sound object-level mappings are rarely perceived within the sound object time scale. Instead, the cumulative effects of such mappings are much more apparent within the meso time scale.

5. DISCUSSION

It is arguable that any audiovisual composition that explores correlations between mediums may be discussed in terms of parameter mappings. As such, it is clear that the field of musical instrument interface design provides a set of terminology relevant to the creation or analysis of audiovisual mapping strategies. Such strategies may be defined as explicit or implicit, static or dynamic and serial or parallel, and may vary in complexity from simple single-level strategies to complex multi-tiered structures. The transparency of audiovisual mappings may be defined as the perceptual correspondence between mapped parameters and it is arguable that simpler mappings will be appear more transparent than those of greater complexity. From a compositional perspective, mapping strategies may exist within the micro, sound object, meso and macro time scales whether explicitly or implicitly defined. On this basis, if an audio grain may be described as a micro-level object it is arguable that a particle may be the visual analogue. Similarly the audio cloud appears to parallel the structure of a particle system within the visual domain; a sound object or meso-level structuring of micro-level particles. On this basis, the term particle may be used to describe a conceptual audiovisual object that exists within the micro time scale. Similarly the term cloud may be used to describe particle structures within either the sound object or meso time scales, based on duration.

Based on observations made throughout the composition process of the stated works numerous assertions may be made with regards to mapping strategies within audiovisual composition. Similar to the observations made by Hunt and Kirk (2000), Neville and Jones (2005) and Dannenberg (2005), mappings of greater complexity appear more successful in terms of sustained interest. Indeed, once perceived the simple parameter mappings within Cirrus soon become predictable, a problem negated slightly by the use of unmapped elements. A similar observation may be made of Photuris. In this case, however, it would appear that it is the linearity of the mappings that leads to potential predictability. Again, the problem is reduced through the use of unmapped elements, but a more dynamic approach to mappings may prove more successful. Based on the success of unmapped elements within the example compositions, it would appear that audiovisual works may benefit from dissonant correspondences between mediums to provide interest and unpredictability.

6. FURTHER WORK

The discussed framework for parameter mapping within audiovisual composition may be extended in many directions. Continuing the conceptual correlation between granular synthesis and visual particle systems, more explicit mapping strategies may be explored within the micro, sound object and meso time scales, as can generative and concurrent mapping systems using an external data set such as behavioural simulation. Further analysis of such compositions will allow the discussion of comparative successes and failures within the context of audiovisual mapping strategies and may identify further techniques relevant to the creation and analysis of audiovisual works.

7. CONCLUSION

This paper has offered that, where a correspondence between mediums is sought, audiovisual composition may be reduced to a series of explicit or implicit maps of varying complexity and transparency, within various compositional time scales. Terminology from the fields of instrument design and algorithmic composition has been offered to provide a framework for the creation of mapping strategies within audiovisual composition. The conceptual objects particle and cloud have also been presented as cross-modal descriptors for micro, sound object and meso-level structures. Observations made throughout the composition of the discussed audiovisual works has identified the importance of mapping complexity within linear composition and has highlighted the relevance of unmapped elements within the creation of audiovisual counterpoint. While a parameter mapped approach to audiovisual composition, may not always be successful, or even desirable, it is clear cross-modal correspondences may be discussed within the framework presented.

8. REFERENCES

Abbado, A. (1988) 'Perceptual Correspondences of Abstract Animation and Synthetic Sound'. *Leonardo, Supplemental Issue,* 1 (, Electronic Art), pp. 3-5.

Alexander, A. & Collins, N. (2008) 'Live Audiovisuals'. *In:* N. Collins & J. d'Escriván eds. *The Cambridge Companion to Electronic Music*.

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.

Bailey, N., Fells, N. & Moody, N. (2007) 'Ashitaka: An Audiovisual Instrument'. *Proceedings of the 2007 Conference on New Interfaces for Musical Expression.* New York, NY, USA

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

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.

Düchting, H. (1997) *Paul Klee: Painting Music.* Prestel.

Edmonds, E., Martin, A. & Pauletto, S. (2004) 'Audio-visual interfaces in digital art'. *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology.* New York, NY, USA

Edmonds, E. & Pauletto, S. (2004) *Audiovisual Discourse in Digital Art.* Los Angeles, California: ACM.

Fels, S., Gadd, A. & Mulder, A. (2002) 'Mapping transparency through metaphor: towards more expressive musical instruments'. *Organised Sound,* 7, pp. 109-126.

Galeyev, B.M. (1976) 'Music-Kinetic Art Medium: On the Work of the Group 'Prometei' (SKB), Kazan, U.S.S.R.'. *Leonardo,* 9 (3), pp. 177-182.

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

Gerhard, D. (1999) 'Audio visualization in phase space'. *Bridges: Mathematical Connections in Art, Music and Science,* , pp. 137–144.

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

Jones, R. & Neville, B. (2005) 'Creating Visual Music in Jitter: Approaches and Techniques'. *Computer Music Journal,* 29

(4), pp. 55-70.

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.* National University of Singapore Singapore, Singapore, pp. 54.

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

Roads, C. (2004) *Microsound.* The MIT Press.

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

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

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

Whitney, J. (1980) *Digital Harmony: On the Complementarity of Music and Visual Art.* Byte Books.

Zilczer, J. (2005)'Music for the Eyes: Abstract Painting and Light Art'. *In: Visual Music: Synaesthesia in Art and Music Since 1900* Thames and Hudson, .

Zilcer, J. (1987) 'Color Music: synaesthesia and nineteenth-century sources for abstract art.'. *Artibus Et Historiae,* 16, pp. 101-126.